Functional investigation of the hbs1l-myb intergenic region on chromosome 6q

Aktuna, Suleyman

Awarding institution:
King's College London

FUNCTIONAL INVESTIGATION OF THE *HBS1L-MYB* INTERGENIC REGION ON CHROMOSOME 6Q

SULEYMAN AKTUNA

Submitted for PhD

Abstract

Fetal haemoglobin levels have a major impact on the severity of sickle cell disease (SCD) and Beta-Thalassaemia. Previous studies by our group shows that three loci – the *HBS1L-MYB* intergenic region (*HMIP*) on chromosome 6q23, *BCL11A* on chromosome 2p16 and the β globin cluster on chromosome 11 accounts for up to 50 % of the variation in HbF levels in non-anaemic Europeans. Recent work by our group indicated the presence of regulatory elements in *HBS1L-MYB* intergenic region (*HMIP*) evidenced by GATA-1 binding coinciding with strong histone acetylation, RNA polymerase II activity, and erythroid specific DNase I hypersensitive sites. Genetic studies and resequencing of two key individuals from the family that led to discovery of 6q QTL revealed new SNPs in the *HBS1L-MYB* intergenic region.

In silico predictions suggested that a couple of these SNPs altered binding of key transcription factors. We proposed that differential binding of these transcription factors mediated through these SNPs underlie the association with HbF control.

Electrophoretic Mobility Shift Assays (EMSAs) and Chromatin immunoprecipitation followed by real time PCR (ChIP-qPCR) and HaploChIP showed that many of the genetically associated variants altered the binding of key transcription factors such as GATA-1, KLF1 and MYB that have crucial roles in globin regulation as well as regulation of erythropoiesis.

We have also compared ex-vivo erythroid cell culture kinetics and expression profiles of individuals who are homozygous for the presence or absence of the allelic variant associated with HbF. *MYB* expression and maturation rate of erythroblasts were noticeably different whereas expression of *HBS1L the* other flanking gene was not different. We could also demonstrate unequal expression of *MYB* in individuals

heterozygous for the *HMIP* variants but the direction of the effect should be further investigated.

These results further support the regulatory role of *HBS1L-MYB* intergenic polymorphisms on HbF and erythropoietic kinetics supporting our hypothesis that the effect of the intergenic SNPs is indirect via the flanking *MYB* gene.

Acknowledgments

The completion of this thesis was made possible by the help and support of many individuals.

My first supervisor Prof. Swee Lay Thein deserves most of the credit for the completion of this thesis. First, I want to thank her for giving me an opportunity for being part of her group. I have to say she has been there for me not only as a supervisor but as a person during rough times and I am extemely grateful for her support. In return, I have given her headaches while reading and editing my thesis but she was ok with that.

I am also thankful for Dr. Steve Best, my second supervisor, for his support during my PhD. His presence made all our lives easier in the lab.

I want to thank the Mol. Haem crew which has been great from the first day and made me a part of the group by asking me to dress up for the fancy dress party on my first day in the lab. There were lots of additions to the group during my period and each and every of them brought positive energy and enthusiasm. They were with me during my wedding as well to share our happiness. I also want to thank to all my friends from JBC which did not run away when I asked them for their blood (You are all safe know).

I want to thank my friends which were a huge part of my life in London and I will miss them all.

I want to thank my mum, my dad and my lovely sister for encouraging me throughout my life. My father's approach to life has been a huge inspiration for meduring my studies.

I have kept the person who is responsible for all this to last. My beautiful wife Acelya, I have followed you to London without hesitation and without you I would not be the man I am today. You have been great during my PhD and I am extremely happy to know that you will be with me during every step of my life.

Table of Contents

List of Figures

List of Tables

1 INTRODUCTION

1.1 Haemoglobin

Haemoglobin is the protein in erythrocytes which acts as a transporter of oxygen and carbon dioxide to all body tissues. Haemoglobin is composed of two alpha-like and two beta-like protein chains, each bound to a hememoiety (Perutz et al 1960).

Different types of haemoglobin are produced at different stages of human development with two developmental switches: Embryonic haemoglobin (Gower 1, $\zeta_2\varepsilon_2$;Gower 2, $\alpha_2\varepsilon_2$ and Portland, $\zeta_2\gamma_2$) to fetal haemoglobin (HbF,$\alpha_2\gamma_2$), and fetal to adult haemoglobin (HbA, $\alpha_2\beta_2$ and HbA_2, $\alpha_2\delta_2$). The first switch from embryonic to fetal haemoglobin occurs in the fetal liverwhile the second, from fetal to adult haemoglobin, occurs just before birth. Haemoglobin A (HbA, $\alpha_2\beta_{2)}$ represents more than 95% of the total haemoglobin in adults and remains the dominant form of haemoglobin throughout adult life (Stamatoyannopoulos 1972 ,Ho and Thein 2000) (Figure 1.1A).

Each of the globin chains are encoded by genes found in two clusters.
The beta-globin gene cluster is located on chromosome 11p; genes in the cluster are arranged in the order in which they are expressed during ontogeny:
the embryonic ε gene, the fetal $^{G}\gamma$, $^{A}\gamma$ genes and the adult δ and β genes, from 5^{l} to 3^{l}, respectively (Figure 1.1 B) (Collins & Weissman 1984, Ho and Thein 2000, Harju and McQueen 2002, Stamatoyannopoulos et al 2005). The major regulatory element of beta-globin gene cluster termed beta-locus control region (beta-LCR) is located further upstream of the genes.

The α-like globin gene cluster is located on the tip of chromosome 16p and genes in the cluster are arranged in the order in which they are expressed during ontogeny: the embryonic α-like globin gene (ζ) and the adult α globin genes (α_1 andα_2) from 5^{l} to 3^{l},

respectively (Figure 1.1 C). The genes are regulated by the upstream regulatory element α-major regulatory element (α-MRE).

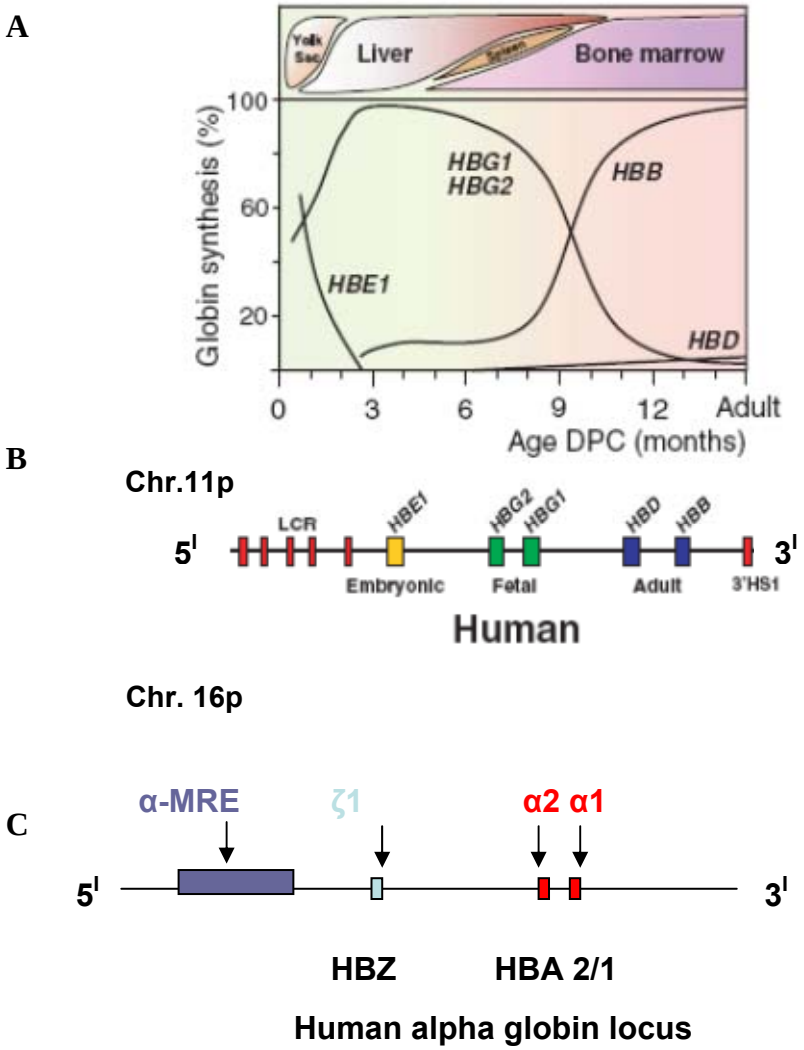


Figure 1.1: Globin expression.
Expression of different forms of β-like globin genes in human development and the developmental switches (A).The β-like globin gene cluster and the locus control region (LCR) in human (B). The α-like globin gene cluster and the α-major regulatory element (α-MRE) in human (C).

This graph is adapted from (Noordermeer & de Laat, 2008).

1.2 Haematopoiesis

Haematopoiesis, the formation of blood cells, consists of a cascade of finely regulated events by which pluripotent haematopoietic stem cells (HSC) differentiate to all the cell types present in blood. These include erythrocytes, lymphocytes, megakaryocytes, granulocytes, and monocytes/macrophages.

Stem cells that have self-renewing capability undergo two sequential differentiating processes in response to specific growth factors, interleukins, and hormones. The first is commitment, by which stem cells lose their self-renewing capability and differentiate to other cells with a more limited differentiating potential (which give rise to only one or sometimes two cell lineages). The second is differentiation, which allows the terminal differentiation of those cells committed to a specific lineage. Differentiation begins after they are committed to either the myeloid or the lymphoid lineage, as a myeloid progenitor cell or lymphoid progenitor cell, respectively.

This process takes place at multiple sites throughout development: Primitive haematopoiesis occurs during the earliest weeks of gestation, in the embryonic yolk sac; definitive haematopoiesis occurs in several locations. The aorta-gonads-mesonephros (AGM) region is the initial site of definitive haematopoiesis followed by the liver and spleen. From around 2 months until 6-7 months of fetal life, the liver and spleen are the major haematopoietic organs, after which, the bone marrow begins to take over the task and eventually becomes the exclusive site of haematopoiesis throughout childhood and adult life (Ogawa 1993, Dzierzak 2005) (Figure 1.2).

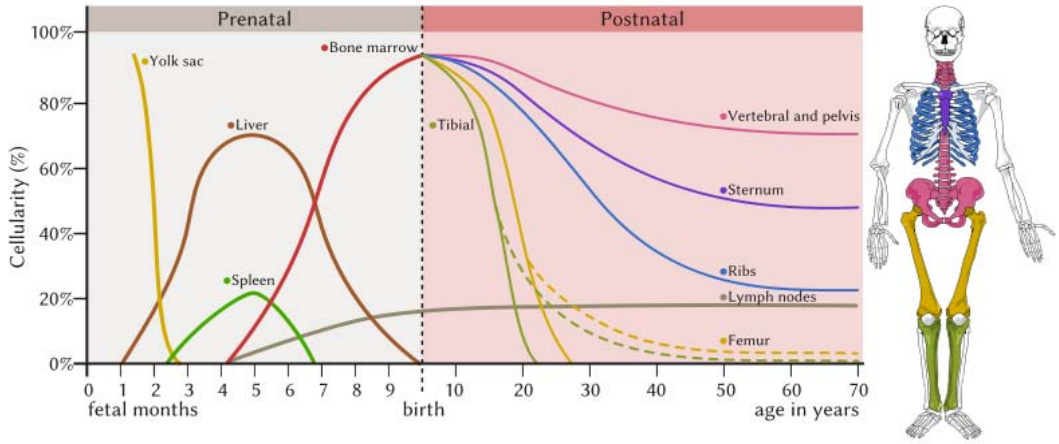


Figure 1.2: Sites of haematopoiesis during human development.
(Image adapted from M.Komorniczak, 2011)

1.3 Erythropoiesis

Erythropoiesis defines the process of differentiation and proliferation from hematopoietic stem cells (HSCs) to mature red blood cells (RBCs).The production of erythrocytes is the largest quantitative output of the haematopoietic system with an estimated production rate of $2x10^{11}$ erythrocytes per day. Erythropoiesis is composed of commitment and differentiation steps that restrict the differentiation potential and the proliferative capacity of the cells as they go through the erythroid-specific program of gene expression. Erythrocytes (red blood cells) are derived from the myeloid progenitor cells. Committed erythroid progenitors can be detected by their ability to form erythroid colonies in *in vitro* methyl cellulose culture, referred to as burst forming unit-erythroid (BFU-e). BFU-e's are the most immature haematopoietic cells that are already committed to the erythroid lineage. BFU-e develops into colony forming units-erythroid (CFU-e). CFU-e will differentiate into the first morphologically identifiable cell of the erythrocyte lineage, the pro-erythroblast, followed by successive differentiation into the basophilic erythroblast, the polychromatophilic erythroblast, and the acidophilic erythroblast, which is the last nucleated cell of the mammalian erythrocyte lineage. In the next stage of development, acidophilic erythroblasts lose the ability to divide, followed by extrusion of the nucleus givingrise to the reticulocyte.

Reticulocytes are released into the blood stream where they mature into erythrocytes (Gregory 1978, Philipsen & Wood, 2009) (Figure 1.3).

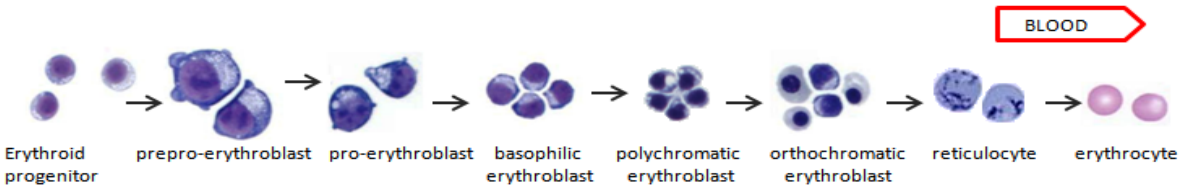


Figure 1.3: Human erythroid differentiation.

Erythroid differentiation of cells grown by using the two-phase liquid culture method.Figure adopted from Philipsen & Wood, 2009.

1.3.1 Regulation of Erythropoiesis

Erythropoiesis must be regulated positively and negatively to ensure a continuous but controlled production of red cells in order to provide oxygen at a physiologic level in peripheral tissues.

A positive regulation is required to ensure a constant production of red blood cells because erythrocytes in circulation have a limited life span of 120 days. Towards the end of this time they are cleared from the circulatory system as new reticulocytes are continuously released into the peripheral circulation. However, a negative control is clearly required at several levels to avoid overproduction of erythroid cells and polycythemia, which may lead to hyper-viscosity and other problems. The process of erythropoiesis is orchestrated by a complex network of microenvironmental and growth factors that promote the survival, proliferation and/or differentiation of erythroid progenitors, and nuclear factors that regulate the transcription of genes involved in the establishment of the erythroid phenotype.

1.3.1.1 Regulation of Erythropoiesis at the Cellular Level

The growth of erythroid progenitor cells is dependent on specific growth factors that act at all stages of erythroid maturation (BFU-E, CFU-E, erythroid precursors) to prevent apoptosis, induce proliferation, and promote or delay/inhibit differentiation. Delay in differentiation may play an important role by allowing self-renewal and expansion of erythroid progenitors before terminal maturation. As a result, production of red cells is increased.

Several growth factors have been described as crucial in the regulation of erythropoiesis, such as interleukin-3 (IL-3), interlukin-6 (IL-6), granulocyte-macrophage colony stimulating factor (GM-CSF), insulin growth factor like 1 (IGF1), and glucocorticoids.IL-3 is a cytokine that promotes cell survival and proliferation by binding to IL-3R expressed on HSCs through to early progenitor cells. GM-CSF is required again from HSC to BFU-e stage and promotes proliferation. IGF1 synergizes with other cytokines to increase proliferation of early erythroid progenitors and may coordinate the production of red cells during development and increase of body mass. Similarly, glucocorticoids may contribute to the increase of erythroid production during stress erythropoiesis, delay differentiation enabling self-renewal and expansion of erythroid progenitors before terminal maturation (Bauer et al 1999; Migliaccio & Papayannopoulou 2001). However, the main factors involved in erythroid development are Stem Cell Factor (SCF) and Erythropoietin (Epo).

At the cellular level, BFU-Es are highly dependent on stem cell factor (SCF) for their growth and differentiation, whereas CFU-Es are highly dependent on erythropoietin (Epo). Genetic and biochemical approaches have demonstrated that SCF and Epo, which act on their respective receptors, Kit and Epo-R, are required for erythroid cell production. Studies have shown that mice with mutations in the SCF gene, or its

receptor gene Kit, develop severe anaemia characterised by depressed erythropoiesis. Epo and Epo-receptor (Epo-R) knockout mice die of failure of fetal liver erythrocyte generation (Figure 1.4).

SCF receptor, c-Kit, is expressed in the majority of CD34+ haematopoietic progenitors and its expression is maintained at high levels during the stages of differentiation from BFU-E to CFU-E. At later stages of differentiation, during the maturation of CFU-E, c-Kit expression progressively declines. Thus, SCF exerts its effects on erythroid progenitors during earlier stages of erythroid differentiation by inducing an increase of cell survival and proliferation of early erythroid progenitors. At the same time, SCF delays erythroid maturation and differentiation to allow self-renewal of early erythroid progenitors and precursors before they reach the stage where cells are no longer able to divide. At the later stages of differentiation, SCF prevents apoptosis and induces proliferation, in synergy with Epo (Broudy et al 1997).

Epo is an essential growth factor in erythropoiesis which is dispensable at the earlier stages of erythropoiesis but its role dramatically increases during later stages. Epo drives the proliferation and expansion of the developing erythroid progenitors from BFU-e stage through to terminal differentiation to haemoglobinized cells. Epo administration dramatically increases the number of CFU-e's and erythroid precursors up to the stage of basophilic erythroblasts, mainly by preventing their apoptosis. At the cellular level, Epo is responsible for the finely tuned homeostatic control of erythrocyte numbers by tissue oxygenation.

Although the daily production of erythrocytes is tightly regulated, loss of blood due to bleeding or haemolysis leads to increased reticulocyte production within a few days. This is achieved by EPO secretionwhich increases the number of CFU-E and erythroid precursors up to the stage of basophilic erythroblasts, mainly by preventing their

apoptosis. As the increased reticulocyte production normalizes circulating erythrocytenumbers, the rate of new reticulocyte formation decreases to avoid reboundpolycythaemia.As the circulating erythrocyte numbers and oxygenation of tissues increase, the production of Epo decreases. Variation of Epo sensitivity at different stages of erythropoiesis is key in the prevention of over production of reticulocytes. . Erythroid progenitors and precursors exhibit a wide range of Epo sensitivities. At low Epo concentration cells with highest Epo sensitivity can prevent apoptosis by EPO administration (Sathyanarayana et al 2008).

Epo interacts with its receptor Epo-R to trigger signaling pathways that regulate erythroid regulation. One of the main signaling molecules activated by Epo binding to Epo-R is STAT5. Activated STAT5 enhance gene activity involved in cell survival, proliferation, and differentiation. Studies have shown that STAT5 plays an important role in maintaining a high erythropoietic rate during fetal development and during stress responses in adult mice. Mutations affecting STAT5 activation affects stress erythropoiesis (Halupa et al 2005).

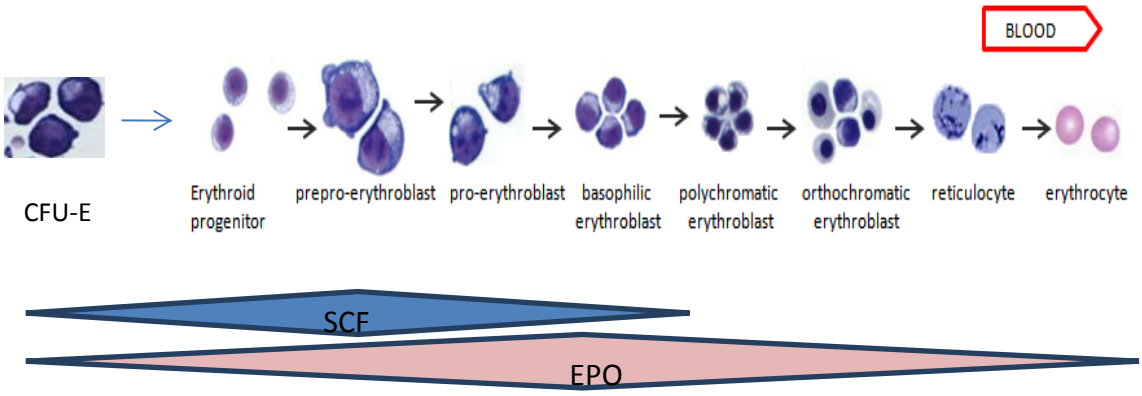


Figure 1.4: Dependence of erythroid progenitor cells to Epo and SCF at different stages of erythropoiesis.
Figure adopted from Philipsen & Wood, 2009.

1.3.1.2 Transcriptional control of Erythropoiesis

At the transcriptional level, the expression pattern of erythroid-specific genes follows a precise timing in the regulation of erythroid differentiation. Studies on the *cis* and *trans*acting factors that regulate the erythroid-specific genes have enlightened us of the complexity of transcriptional regulation of erythropoiesis.

The action of critical transcription factors determines the selective expression of cell type-specific genes. These transcription factors can be either expressed constitutively or induced at a certain stage of cell differentiation (Orkin et al 1992; Mikkola and Orkin, 2006).

Studies suggested that the GATA family proteins have crucial roles in erythropoietic regulation. Presence of GATA motifs is crucial for the lineage specific expression of erythroid specific genes. Transcription factors GATA-1 and GATA-2 have crucial roles, both in erythrocytic determination and in terminal differentiation of erythropoiesis. During erythropoiesis, there is sequential but overlapping expression of GATA-1 and GATA-2. GATA-2 is expressed early in the erythroid lineage and its expression decreases with the increase of GATA-1 expression. GATA-2 is involved in the cellular proliferation of non-committed and committed erythroid progenitors, whereas GATA-1 is critical for terminal maturation of erythroid cells. At the transcriptional level, GATA-2 inhibits transcription of the GATA-1 gene and GATA-1 has the same effect on the GATA-2 gene. The regulatory network between GATA-1 and GATA-2 during erythropoiesis results in the sequential substitution of GATA-2 by GATA-1 in the regulatory regions of numerous erythroid specific genes (Tsai et al 1994; Tsai and Orkin, 1997; Ferreira et al 2005).

GATA-1 transcriptional activity is also dependent on protein-protein interactions. Many proteins have been shown to interact physically with GATA-1 including Friend of GATA-1 (FOG-1)**,** LIM domain only 2 (LMO- 2), T-cell acute lymphocyticleukaemia 1 **(**TAL-1), growth factor independent 1B transcription repressor (Gfi-1b) and Spleen

proviral integration oncogene (SP1). FOG-1, a multitype zinc finger protein, specifically interacts with the amino zinc finger of GATA-1 and is especially important for GATA-1 activity. Studies suggested that function of GATA-1 in erythropoiesis is linked to FOG-1 and this was evidenced by the observation that a single amino-acid change in the GATA-1 amino zinc finger, which abolishes interaction with FOG-1, is lethal in mice due to severe anemia (Cantor et al 2005).

The Myeloblastosis viral oncogene homolog (MYB) transcription factor plays crucial roles in regulation of proliferation and cell lineage commitment of erythroid progenitor cells. It is required for maintenance of proliferation by exerting a direct role in cell cycle control. MYB is highly expressed in immature haematopoietic cells and is down-regulated as the cells differentiate. Expression of MYB blocks differentiation in neuroblastoma cell lines, suggesting down-regulation of MYB is also required for terminal differentiation (Westin et al1982, Ohet al 1999) (Figure 1.5).

Another key factor,Kruppel-like factor 1(KLF1), is required for the last steps of erythroid differentiation. It directly regulates genes involved in haemoglobin metabolism and membranestability. KLF1 knockout mice die from anemia, supporting the role of KLF1 during erythropoiesis (Bieker et al 2005).

1.4 Control of Globin gene expression

Expression of the globin genes are strictly regulated at all levels, so that synthesis of the alpha-like and beta like globin chains are balanced throughout human development. Imbalanced globin chain synthesis caused by the reduction of one of the globin chains results in thalassaemias (Fritsch et al 1980).

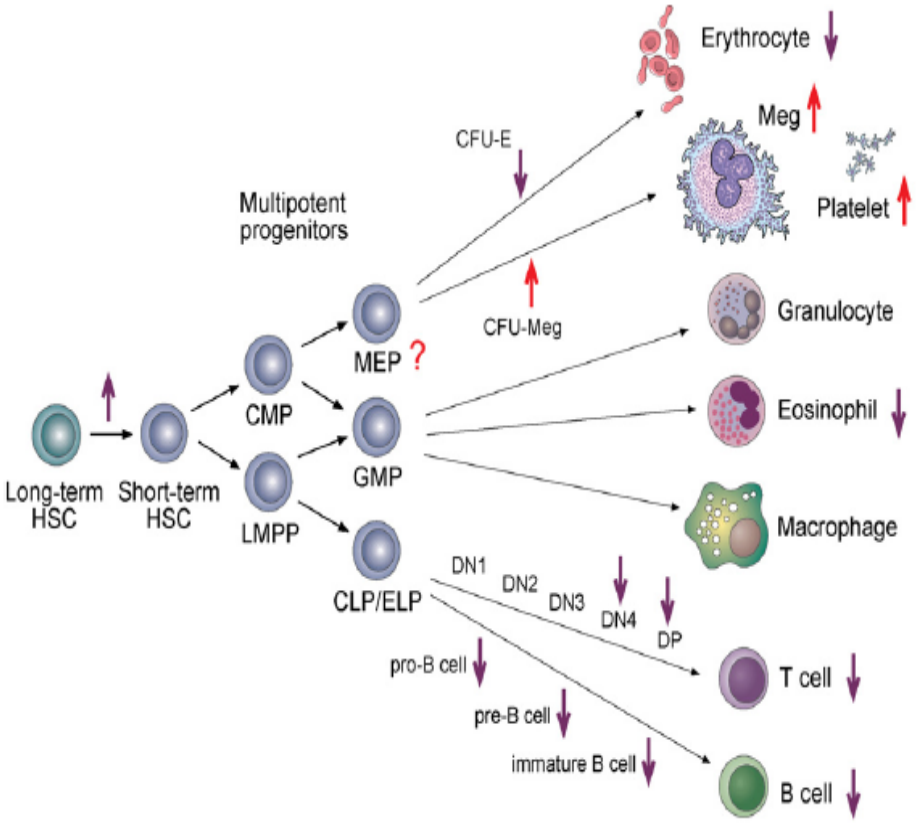


Figure1.5: Diagram of hematopoiesis, illustrating the multiple defects observed in mouse models of MYB deficiency.

(↓): A decrease in the frequency of the designated cell type/maturation stage in the absence of MYB.

(↑): Cell populations whose frequency is increased by the loss of MYB function.

(Figure adopted from K.T. Greig et al / Seminars in Immunology 2008).

Abbreviations: LMPP, lymphoid-primed multipotent progenitor; CMP, common myeloid progenitor; GMP, granulocyte–macrophage progenitor; CLP/ELP, common lymphoid progenitor/early lymphoid progenitor; Mep, Myeloid erythroid progenitor; Meg, megakaryocyte.

The expression pattern of globin genes depends on local cis-acting sequences at the promoter sequences. The globin gene promoters share a number of characteristic

sequences at specific distances from the transcriptional start site (TSS). These sequences include TATA, CCAAT and CACCC which are situated 30, 70-78, and 80-140 bp upstream of the TSS, respectively. The TATA and CCAAT sequences are found in many eukaryotic promoters while the CACCC sequence exists predominantly in erythroid specific genes. Importance of these sequences for normal globin gene expression has been demonstrated by the presence of natural mutations at these sites that lead to down regulation of the globin genes causing thalassaemia. Other regulatory sequences that control the globin genes include the enhancer and silencer elements within the *HBB* complex that interact with protein complexes to influence promoter activity and transcription of target genes (Forrester et al 1987; Grosveld et al 1987; Higgs et al 1990). Remote cis-acting sequences, such as the locus control region (LCR) of the β-globin gene cluster and the major regulatory element (MRE) of the α-globin gene cluster are also crucial for the expression ofthe globin genes (Higgs et al 1990).

The function of a powerful upstream enhancer of the β-globin loci, the locus control region (LCR), was identified as essential for high level expression of these genes (Tuan et al 1985; Forrester et al 1986; Grosveld et al 1987; Bender et al 2000) Molecular experiments such as DNAse Hypersensitivity assays enlightened us with the active parts of the β-globin locus. Genes that have an 'open' chromatin structure (euchromatin) are transcriptionally active, whereas transcriptionally silent genes have a 'closed' chromatin structure (heterochromatin). Open chromatin is hypersensitive to nuclease digestion while closed chromatin is resistant to DNase I digestion. Thus, transcriptionally active sites can be identified by their hypersensitivity to DNase I. In 1980s Tuan et al and Groudine et al discovered the 5 hypersensitive sites that are located upstream of the ε-globin gene (Figure 1.6). These sites are phylogenetically conserved sequences and they play a crucial role in the regulation of the beta globin gene locus (Forrester et al 1986, Stamatoyannopoulos et al 2005). There are 2 more

hypersensitive sites that are positioned further upstream but only hypersensitive sites 1 to 5 were shown to be erythroid specific (Stamatoyannopoulos et al 2005).

Figure 1.6: 5 hypersensitive sites located upstream of the ε-globin gene.
Figure adapted from Noordermeer & de Laat, 2008.

The LCR spans a 16kb region which starts 5kb upstream of ε-globin gene. It has important functions, such as enhancement of transcription, remodeling the chromosome structure through histone modifications, acting as an insulator to regulate enhancer promoter interactions and enabling the correct positioning of the beta globin locus in the nucleus (Stamatoyannopoulos et al 2005, Ragoczy et al 2006).

Various studies supported the importance of the LCR for globin regulation; as evidenced by results that deletions of the β-LCR in transgenic mice leads to reduction of globin gene expression by 100-fold. Another study showed that deletion of individual HS sites in the β-LCR disrupts the activation of globin genes proving that an intact LCR is necessary for activation of the globin genes (Milot et al 1996, Bender et al 2000).

At the transcriptional level, specific sequences in the promoters compete with other promoters to interact with the Locus Control Region (LCR) through binding of transcription factors. Binding of these transcription factors is regulated through remodeling of the chromatin by various histone modifications that enables activation of different parts of the globin locus during development. Acetylation and methylation are the histone modification observed at the beta globin locus. Acetylation mainly harbors

an open state of chromatin thus acts as an activator. On the other hand methylation may act both as an activator and repressor depending on the site that is methylated. According to Berger et al methylation of lysines at positions 4, 36 and 79 on the third histone (H3) correlates with activation while methylation of lysines at positions 9 and 27 mainly leads to repression (Berger et al 2007).

1.4.1 Globin gene activation by the β-LCR

Several models have been proposed to explain how the β-LCR functions to activate the downstream globin genes and of these, the looping model is the most widely accepted (Figure 1.7). This model suggests that the HS sites of the β-LCR fold into a holocomplex, bringing the LCR closer to the appropriate gene and thereby the transcription factors bound to the LCR are delivered to the gene locus to interact with the basal transcriptional apparatus and activate globin gene expression (Ellis et al 1996; Fraser et al 1993; Stamatoyannopoulos 2005; Wijgerde et al 1995).

Studies have demonstrated the interaction of the LCR with downstream globin gene promoters via the looping model. Carter et al (2002) used an RNA FISH method called RNA TRAP to tag and recover chromatin near the murine β-globin gene whichrecovered HS1, HS2 and HS3 together with the β-globin gene. This demonstrated that the HS sites and β-globin gene were in close proximity. A further study used chromatin conformation capture (3C) to measure the proximity of LCR sequences and active globin genes in erythroid cells. Clustering of active regulatory elements at these sequences was referred to as an active chromatin hub (ACH) (Tolhuis et al 2002).

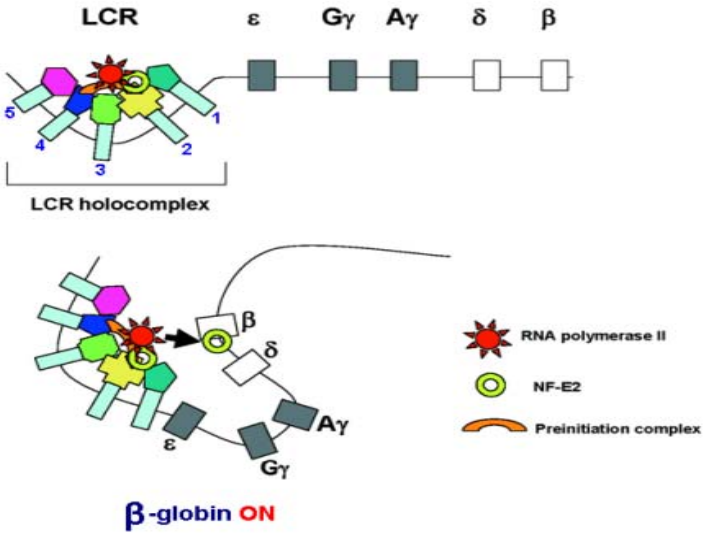


Figure 1.7: The looping model of gene activation at the β-globin locus.
The LCR holocomplex is formed with transcription factors and other transcriptional apparatus. The holocomplex loops over the globin genes, bringing the transcriptional apparatus closer to the globin genes for transcription of the preferential gene (the β-globin gene in this example). Image adapted from Vieira et al 2004.

1.4.1.1 Gene switching mechanisms

Two mechanisms - autonomous silencing and gene competition - were proposed in order to explain how certain genes are switched off while others remain turned on. Gene competition was first investigated in transgenic mice in which the β-LCR was linked to either theγ-globin (μLCR ^{A}γ) or the β-globin (μLCR β) gene, alone. In both cases, the genes failed to show developmental expression, i.e. the β-globin gene was expressed in cells throughout all developmental stages including the embryonic cells.In other words, β-globin gene expression throughout development was as it would be in adult cells. The γ-globin genes were also expressed in adult cells (Enver et al 1990; Stamatoyannopoulos 2005).

When the LCR, γ and β genes were linked together in the same construct (μLCR ^{A}γψβδβ), developmental expression of the genes followed. The γ-globin gene was predominantly expressed in embryonic erythrocytes and theβ-globin gene was highly expressed in adult mice erythrocytes. γ and β-globin genes were expressed similarly in fetal liver erythroblasts, similar to the human perinatal stage (Enver et al 1990). The investigators concluded that the γ and β genes compete to interact with the LCR (Enver et al1990). During fetal life, proximity to the β-LCR leads to preferential expression of the γ-globin gene. During adult life the γ-globin gene is autonomously silenced and the LCR interacts with the next downstream gene i.e. β-globin.

Similar experiments in transgenic mice showed that mice carrying the embryonic ε-globin gene do not express the gene at all, but when the ε-globin gene is linked to the LCR (μLCR ε) the gene is expressed at high levels, indicating the ε-gene expression is dependent on the LCR. At the same time, ε-globin gene expression was not observed in yolk sac erythropoiesis, suggesting all the components needed to turn off a gene are contained within the gene or adjacent sequence. This concept is called 'autonomous silencing' (Raich et al 1990). Transcription factors bind on silencing sequences at the proximal and distal ε-globin gene promoters to form a silencing complex which turns off ε-globin gene expression by disrupting interactions between the gene and LCR (Stamatoyannopoulos 2005). Autonomous silencing is thought to be the principal mechanism for turning off genes, though gene competition is also contributing.

1.4.2 Transcription Factors

Transcription factors play a crucial role in globin switching. They bind to specific recognition sequences or binding motifs and help in the activation or repression of genes acting as *trans* regulators. Several transcription factor-binding motifs have been characterized within the β-globin locus and are essential for the formation of the active

locus and activation and control of individual genes. The core regions of the hypersensitive sites in the LCR are particularly rich in such motifs and include binding sites for the erythroid specific transcription factors KLF1, GATA-1, TAL-1 and NF-E2 as well as ubiquitous factors such as SP1 (Goodwin et al 2001; Ho & Thein 2000). Table 1.1 summarizes key functions of critical transcription factors.

Transcription Factor	**DNA Binding Motif**	**Function**
GATA binding protein 1 **(GATA-1)***	(T/A)GATA(A/G)	GATA-1 is the master regulator of erythropoiesis and key factor for globin switching
Kruppel-like factor 1 **(KLF1)***	CACCC	KLF1 has crucial roles in cell proliferation, differentiation and survival KLF1 acts as an activator of BCL11A which represses expression of γ-globin
Myeloblastosis viral oncogene homolog **(MYB)**	YAAC(G/T)G	MYB have crucial roles in regulation of proliferation and cell lineage commitment of erythroid progenitor cells MYB influences HbF levels via altering erythroid kinetics
T-cell acute lymphocytic leukaemia 1 **(TAL1 / SCL)**	CAGATG	Member of pentameric complex that regulates erythroid differentiation and gene expression. SCL plays an essential role in early haematopoiesis.
LIM domain only 2 **(LMO2)**	CAGGTG	It acts as an adaptor protein that mediates interactions between different classes of transcription factors and their cofactors.

Nuclear factor (erythroid-derived 2) **(NF-E2)***	TCA(T/C)	NF-E2 binding is important for transcriptional activation and formation of HS sites in the LCR
B-cell CLL/lymphoma 11A **(BCl11A)**	GGCCGG	BCL11A acts as a repressor of γ-globin It is also essential for normal B and T cell development and lymphogenesis.
Friend of GATA-1 **(FOG1)***	Cofactor	FOG-1 plays an essential role in erythropoiesis and megakaryopoiesis Essential cofactor that acts via the formation of a heterodimer with GATA-1
SRY (sex determining region Y)-box 6 **SOX6**	5′-WWCAAW-3′ (W=A/T)	SOX6 play a role in the silencing of the mouse embryonic globin genes Interacts with BCL11A to regulate globin genes
Direct repeat erythroid-definitive **(DRED)*** / Nuclear receptor subfamily 2 **TR2/TR4***	CAAT	DRED binds with high affinity to DR1 sites in the human embryonic (ε-) and fetal (γ-) globin gene promoters TR2/TR4 forms the core of a large DRED complex that represses embryonic and fetal globin transcription in definitive erythroid cells

Table 1-1: Key transcription factors in haematopoiesis and erythropoiesis.

(* = erythroid specific transcription factors)

1.4.2.1 GATA-1

GATA binding transcription factor 1(GATA-1) is required for globin gene switching and erythroid cell maturation and mainly expressed in erythroid and megakaryocytic lineages .It is a member of the GATA zinc finger family of transcription factors which bind to nucleic acid consensus sequence (T/A)GATA(A/G) (Evans & Felsenfeld 1989; Tsai et al 1989).

GATA-1 binds to sites in the globin gene promoters and HS site cores of HS1-5 which contain the GATA-1 recognition sequence (Orkin 1992). GATA-1 can function both as a repressor or an activator of gene expression and the function is partly dependent on its interactions with other proteins.

GATA-1 can interact with itself as well as other transcription factors including KLF1 and SP1 and can form large multiprotein complexes with TAL1, E47, LDB1, and LMO2. Strong positive correlation between occupancy of DNA by GATA1 and TAL1 was also shown (Crossley et al 1995; Merika & Orkin, 1995; Wadman et al 1997; Morceau et al 2004; Cheng et al 2009). Thus, GATA-1 is involved in many aspects of globin locus activity. It is also involved in the establishment of the active chromatin domain in the locus and the formation of hypersensitive sites (Orkin et al 1992; Pomeratz et al 1998; Vakoc et al 2005; Layonet al 2007).

1.4.2.2 KLF1

Krüppel-like factor 1 or KLF1 (previously known as erythroid Krüppel-like factor, EKLF) was identified by subtractive hybridization and found to be an erythroid cell-specific transcription factor, homologous to the Krüppel family of transcription factors which have roles in cell proliferation, differentiation and survival (Miller & Bieker, 1993). KLF1 contains three zinc fingers at the C-terminus which bind to a CACCC sequence. CACCC sequences are repeated in erythroid enhancers and promoters, including the

β-globin gene promoter (Miller & Bieker, 1993). This sequence is noted as a site of point mutations that give rise to β-thalassaemia. It is reported by Feng et al(1994) that KLF1 is unable to transactivate in the presence of these point mutations due to a decrease in binding affinity for these target sites. The KLF1 protein binds to the β-globin promoter with 8 fold higher affinity than to the γ-globin promoter (Donze et al 1995). These studies suggested KLF1 is acting as a stage-specific, β-globin-specific transcription factor, and most likely to be important in the fetal to adult (γ to β) haemoglobin switch.

Further studies in mice revealed that *Klf1*$^{-/-}$ mice die around day 14-15 of gestation due to anaemia caused by the failure to express β-globin. Moreover, the embryos showed features of β-globin deficiency as found in β-thalassaemia. *Klf1*$^{-/-}$ embryos appeared normal during the embryonic yolk sac stage of haematopoiesis, but became fatally anaemic during early fetal life, at the precise time of the switch from embryonic to fetal-liver erythropoiesis (Nuez et al1995 & Perkins et al 1995).The activation of the different β-globin like genes is mediated in part through the β-globin LCR region. KLF1 binding enhances the interaction between the β-globin promoter and the LCR and so the interaction between the γ-globin promoter and LCR decreases (Miller & Bieker 1993, Donze et al 1995).

The human β-globin locus was studied in *Klf*knockout/human β-locus transgenic mice (Wijgerde et al 1996). The ε and γ*globin* genes were expressed normally in *Klf1*$^{-/-}$foetuses, with a complete lack of β-globin expression. In *Klf1*$^{+/-}$/β-globin transgenic mice, there was a shift in the β toγ ratio caused by an increase in the number of actively transcribed γ genes and a decrease in transcribed β genes. The authors proposed that the reduction in klf1 in the K*lf1*$^{+/-}$mice reduced the time the LCR was in contact with the β-globin gene. Because KLF1 does not directly bind γ-globin promoter, an increase in γ-globin is due to the shift in interaction of the LCR from

β-globin gene promoter toγ-globin gene promoter (Wijgerde et al1996). This was accompanied by changes in chromatin structure at the β-globin gene promoter, and HS3 of the β-LCR.

Following this work, KLF1 was found to be required for the activity of 5'HS3 of the β-globin LCR by binding directly to the core fragment within HS3 (Gilemans et al 1998). Increasing levels of KLF1 lead to changes in the balance from γ to β-globin gene expression, which results in an earlier switch of the globin genes and the amount of KLF1 also influences the rate of the switching process (Tewari et al 1998).

These studies, collectively demonstrate the importance of KLF1 in the γ to β globin switch and LCR-β-globin gene interactions.

Recently it has been proposed that KLF1 has a direct role in *BCL11A* activation. Borg et al (2010) mapped the cause of HbF persistence in a Maltese family to chromosome 19 and found a heterozygous nonsense*KLF1* mutation in family members with high levels of HbF. *BCL11A* expression was downregulated in the erythroid progenitor cells of these family members; moreover, γ-globin expression was upregulated. Further, knockdown of KLF1 in primary human erythroid progenitor cells resulted in an increase in γ-globinand a decrease in BCL11A. KLF1binding was shown at the *BCL11A* promoter, demonstrating the direct role of KLF1 on *BCL11A.* These findings conclude that HbF persistence here is caused by *KLF1* haploinsufficiency. KLF1 activation of *BCL11A* was confirmed by others (Zhou et al 2010), proposing a new model for globin regulation (Figure 1.7).

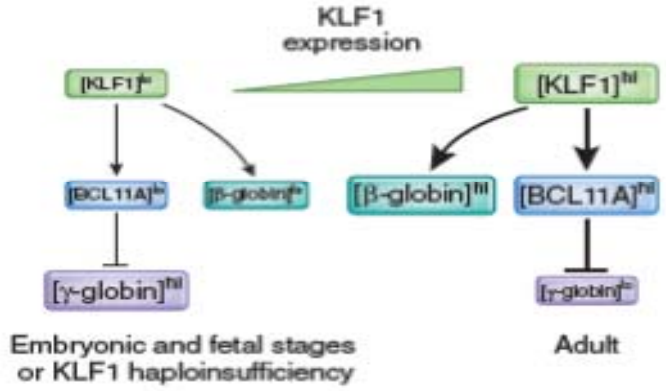


Figure 1.8: Regulation of globin genes by KLF1.
KLF1 acts as an activator of BCL11A which represses expression of γ-globin. Hence there is a negative correlation between the expression levels of KLF1 and γ-globin. Figure was adopted from Bieker et al Nature 2010.

1.4.2.3 MYB

The *MYB* gene (also known as *c-Myb*) was first identified as the cellular homologue of *v-Myb*, an oncogene found in two avian retroviruses that induce leukemia.

The *MYB* gene is highly conserved through evolution (Lipsick, 1996); the major product of the human gene is a 75kDa nuclear protein of 636 amino acids. The MYB protein is expressed in most haematopoietic tissues and is highly expressed in immature haematopoietic cells (Gonda et al 1982, Westin et al 1982).

The MYB protein (75 kDa) has an isoform of 89 kDa (cMYB-9A) caused by a splice variant which leads to an insertion of 363 base pairs (bp) between exon 9 and 10 (exon 9A) The p89 form of MYB displays higher transactivation activity, however little is known about its functional relevance. Jiang et al confirmed that the pattern of *MYB-9A* expression was similar to that of the total *MYB* expression in erythroid progenitors.

Following the molecular cloning of *Myb*, numerous studies performed in cell lines implicated this gene in the proliferation, survival and differentiation of hematopoietic cells. The earliest of these was a null allele (*MYB*−) that established *MYB* as an essential gene for definitive hematopoiesis. Further evidence that MYB also plays a physiological role came from studies of mice that lack MYB, which die embryonically with failures of erythroid and myeloid development. Studies also suggested that lack of MYB did not alter the generation of HSCs but failure of erythropoiesis was due to a defect in expansion and/or differentiation of these cells (Mucenski et al 1991, Sumner et al 2000).

Efforts were made to identify MYB target genes, by measuring differentially expressed genes between $MYB^{-/-}$ null mutants and wild-type cells. The GATA-1 transcription factor was found to be down-regulated in the $MYB^{-/-}$ cells, giving clues to the involvement of MYB in regulation of haematopoiesis. Interaction of MYB and GATA-1 during haematopoiesis was speculated following this result. In 1996 Briegel et al suggested an antagonistic relationship between GATA-1 and MYB. This was later supported by studies showing that GATA-1 and MYB crosstalk during red blood cell differentiation through GATA-1 binding sites in the *MYB* promoter (Lin et al 1996, Bartunek et al 2003).

Until recently the direct role of MYB on the regulation of the γ-globin gene was elusive. It was proposed that MYB influences HbF and F cells levels via altering erythroid kinetics (Jiang et al 2006). It is evident that MYB is involved in cell proliferation and differentiation; low *MYB* levels correlate with accelerated erythroid maturation. Another mechanism whereby MYB can influence HbF levels via KLF1 activation of *BCL11A* is proposed .The study suggests a coorporation between MYB and KLF1; in which MYB supports erythropoiesis through the transactivation of KLF1 and LMO2 expression (Bianchi et al 2010).

1.4.2.4 BCL11A

BCL11A is a highly conserved zinc finger gene; it is expressed in hematopoietic progenitors, down-regulated during myeloid differentiation (Saiki et al2000) and is essential for normal B and T cell development (Liu et al2003). BCL11A binds to GC-rich motifs and mediates transcriptional repression (Senawong et al2005, Sankaran et al2008). It is also involved in lymphoid malignancies through translocation or amplification events (Fell et al1986, Richardson et al 1992, Satterwhite et al 2001). *BCL11A* mutations have also been associated with B-cell malignancies (Liu et al 2006).

It has been shown that BCL11A acts as a repressor of HbF production. Knockdown ofBCL11Ain primary human erythroid cells led to a dramatic elevation of the levels of γ-globin and HbF (Sankaran et al 2008). However, this did not affect overall erythroid differentiation. BCL11A has also been shown to bind to discrete sites in the β-globin gene cluster, indicating a direct role of *BCL11A* in globin gene regulation (Sankaran et al 2008, Jawaid et al 2010). Studies using a human β-globin locus YAC transgene and a *Bcl11a* $^{-/-}$ mouse, showed that in the absence of Bcl11a, developmental silencing of the human γ-globin genes is impaired in the definitive erythroid lineage. Intermediate silencing in a *Bcl11a*$^{+/-}$ mouse suggests that not only is BCL11A a developmental-stage specific repressor, but the effect is quantitative (Sankaran et al2009). Also, differences in *BCL11A* expression patterns between mouse and humans may be responsible for their divergent expression of β-like globin genes (Sankaran et al2009).

At least four *BCL11A* transcripts exist, giving rise to different protein isoforms: extra-long (XL; 5.9kb, 125kD), long (L; 3.8kb, 100kD), short (S; 2.4kb, 35kD) and extra short (XS; 1.5kb, 25kD). Exons 1& 2 are common to all four isoforms, exon 3 & 4 to isoforms XL, L and S. Isoforms L and S share exon 5 and isoform XS contains exon XS (Liu et al 2006) (Figure 1.9).

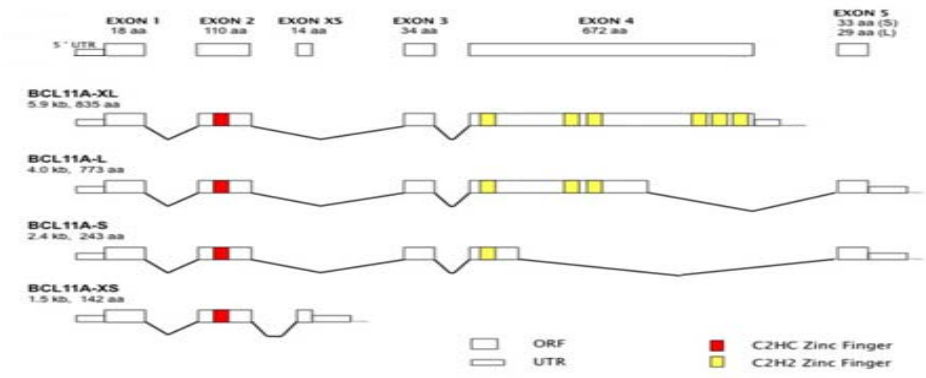


Figure 1.9: Four major BCL11A isoforms, XL, L, S and XS
Image adapted from Liu et al 2006. The role of these different isoforms is largely unclear. Studies have shown that the shorter isoforms appear in primitive erythroid cells, whereas the larger forms appear in later developmental stages of erythroid maturation, when γ-globin expression is reduced (Sankaran et al2008).

1.4.2.5 NF-E2

NF-E2 is an erythroid-specific DNA-binding protein that recognizes motifs (GCTGA (G/C) TCA) in the 5'HS2 of the β-globin LCR (Talbot & Grosveld 1991, Stamatoyannopoulos et al 1995).

NF-E2 binding sites are important for transcriptional activation and formation of HS sites in the LCR (Forsberg et al 2000; Ney et al 1990). Experimental evidence also shows NF-E2 binding sites are important for chromatin remodelling activity and necessary for ε-globin gene expression and formation of HS2 (Gong et al 1996).

1.4.2.6 SCL (TAL1)

T-cell acute lymphocytic leukaemia protein 1 (TAL 1 or SCL), a member of the basic helix-loop-helix (bHLH) family of transcription factors, was first identified by its involvement in recurrent chromosomal translocations in patients with T-cell acute lymphoblastic leukaemia (ALL) (Begley et al 1989).

At the molecular level, TAL-1 acts through both DNA-binding dependent and independent mechanisms and can activate or repress transcription. SCL heterodimerizes with E proteins, such as E2A, and binds to E-boxes (CANNTG).These sequences are found in cis-acting regulatory elements of erythroid-specific genes in conjunction with GATA-binding motifs. The orientation of the E box (CANNTG or CAGATG) and presence of GATA-1 motif are important for enhancing TAL1/SCL binding affinity.

Studies demonstrated that a pentameric transcription factor complex is formed between SCL, LM02, GATA-1, E47 and LDB1 that regulates erythroid differentiation and gene expression. It was also suggested that GATA1, SCL, and KLF1 in combination, delineate active chromatin and enhancer states (Wadman et al 1997, Tallack et al 2010) (Figure 1.10).

At the cellular level, studies demonstrated that $SCL^{-/-}$ ES cells fail to contribute to all hematopoietic lineages in adult chimeric mice and $Scl^{-/-}$ mice die during embryogenesis with a complete absence of yolk sac blood, indicating that SCL plays an essential role in early haematopoiesis (Porcher et al 1996;).

1.4.2.7 LMO2 (RBTN2)

LMO2 (RBTN2) is a LIM domain-containing protein that has been implicated in the pathogenesis of certain T cell ALLs. Studies suggested that SCL and LMO2 physically interact with high affinity evidenced by the similarities in the loss of function phenotype between these transcription factors (Warren et al 1994)

LMO2 was a part of pentameric complex formed between SCL/E2A, LM02, GATA-1, and another LIM-containing protein, LDB1, and it was suggested that its function is to act as a bridge together with LDB1 between SCL/E2A heterodimer bound to an E-box sequence and GATA-1 bound to a GATA element (Wadman et al 1997).

1.4.2.8 LDB1

LIM domain-binding protein 1 (LDB1) is a nuclear protein that contains an N-terminal dimerization domain and a C-terminal LIM interaction domain. LDB1 binds to LIM-homeodomain (LIM-HD) and LIM-only (LMO) proteins. It acts as an adaptor protein that mediates interactions between different classes of transcription factors and their cofactors. LDB1 have been implicated in long-range gene regulation and in regulation of erythropoiesis as a partner of GATA-1 (Cantor and Orkin, 2005, Tallack et al 2010). Understanding the control of globin genes is crucial in the struggle to find new targets to increase the fetal haemoglobinlevels which have ameliorating effects on the haemoglobinopathies.

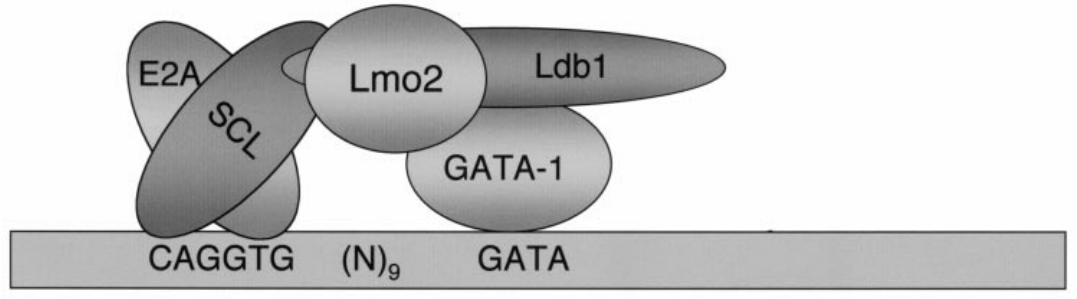


Figure 1.10: Pentameric complex formed between SCL, E2A, Lmo2, Ldb1 and GATA-1.
Schematic representation of cis-acting regulatory DNA from a hypothetical erythroid target gene containing an E-box motif (CAGGTG) and GATA-binding motif (GATA) separated by nine nucleotides. SCL and its heterodimeric partner, E2A, are shown interacting with an E-box sequence and GATA-1 interacting with a GATA motif. These complexes are connected via interactions with LMO2 and Ldb1(Figure adapted from Wadman et al (1997).

1.5 Haemoglobinopathies

Disorders of haemoglobin results from mutations affecting either the structure of haemoglobin (giving rise to haemoglobin variants) or quantitative reduction of the globin chain synthesis (giving rise to thalassaemia) (Weatherall & Clegg, 2001 ;Thein& Clark 2004).

A reduction in alpha globin production causes α-thalassaemia, the most severe form being Bart's hydrops fetalis (--/--) when there is a complete absence of alpha globin resulting in death in-utero. A-thalassaemia carriers (αα/--, α-/ α- and αα/ α-) are normally asymptomatic; individuals with HbH (--/α-) have transfusion-independent anaemia (Higgs, 1993). Beta-thalassaemia and Sickle Cell Disease(SCD), which become manifested soon after birth, are severe conditions that cause a great public health burden particularly among African, Asian and Mediterranean populations(Weatherall et al 2001).

An estimated 300,000 new-borns with these disorders are born worldwide each year, the majority of which have SCD. These are prevalent in tropical and sub-tropical regions where carriers have a survival advantage to *Falciparum* malaria, but these disorders have spread to most countries by population migration. β-thalassaemia is most common in Asia, the Mediterranean and Middle East, and SCD predominates in Africa, the Mediterranean and parts of India (World Health Organization, fact sheet No.308, Jan 2011).

1.5.1 Beta-thalassaemia

Beta-thalassaemia is characterised by the quantitative deficiency of functional β-globin chains, the majority being caused by mutations in the β-globin gene or within its

flanking sequences. The continuing production of α-globin chains leads to an excess of α-chains that aggregates and form inclusion bodies. The inclusion bodies cause damage to the cell membranes and premature death of erythroid precursors in the bone marrow i.e. ineffective erythropoiesis. Cells that enter the microcirculation containing inclusion bodies suffer physical damage with haemolysis as a consequence. Haemolysis is further caused by an increase in reactive oxygen species that damage cell membranes. Ineffective erythropoiesis and haemolysis causes a reduction of erythrocytes and consequently anaemia .Erythropoiesis is thus expanded to compensate for the lack of functional red blood cells and causes hyperplastic marrow; an increased proliferation of progenitors in the bone marrow (Weatherall et al 1998; Thein et al 1998; Schier et al 2002).

The clinical complications of β-thalassaemia result from the severe anaemia and ineffective erythropoiesis which are directly related to the degree of globin chain imbalance. A large part of the variation is caused by variations of the β-thalassaemia alleles but co-existing mutations or variations affecting the levels of α and γ-globin are important contributors. Thus, co-inheritance of α-thalassaemia which reduces the α/β globin chain imbalance produces a milder condition. A persistent increased production of γ-globin chain reduces chain imbalance by combining with the excess α-globin to form HbF (α2γ2), thus ameliorating the severity of β-thalassaemia (Bank et al 2005).

1.5.2 Alpha-thalassaemia

Alpha-thalassaemia is characterised by the deficiency of functional α-globin chains caused mainly by deletions that removes one or both of the α-globin genes. As a consequence, there is an excess of β-chains in adult life, and γ-chains in fetal life but these are capable of forming soluble tetramers. The solubility of these tetramers makes α- thalassaemia a less severe disease. However these tetramers are unstable and

cause haemolysis in the peripheral circulation and a variety of clinical manifestations (Higgs et al 1993; Schrier et al 2002).

1.5.3 Sickle Cell Disease

Sickle cell disease results from the presence of sickle haemoglobin (HbS) which is caused by a single point mutation in the sixth codon of the β-globin gene (CAG→CTG) substituting glutamic acid for valine. HbS ($\alpha_2\beta^s_2$) polymerises in its deoxygenated form through hydrophobic interaction between β^s-chains causing a rigid sickle-like shape of erythrocytes that are trapped in microcirculation causing organ damage. Sickled cells are also highly fragile with a shortened life span, leading to haemolytic anaemia.

The potential and rate of polymerisation is directly proportional to the concentration of HbS in the cell. The presence of other haemoglobins in a cell reduces HbS concentrations and is therefore a limiting factor on polymerisation. Individuals heterozygous for HbS (HbAS; sickle trait) are generally without symptoms as the HbS concentration is below the threshold concentration for polymerisation. Likewise, expression of HbF limits the extent of polymerisation. In addition to diluting the concentration of HbS in the cell, γ-globin has a direct inhibitory effect on polymerisation as mixed tetramers containing a γ-chain ($\alpha_2\gamma\beta^s$) do not participate in polymerisation. Even modest changes in HbF expression can therefore have profound effects on clinical disease (Pauling et al 1949; Noguchi et al 1993; Stuart & Nagel, 2004). Hence, the regulation and control of fetal haemoglobin is deeply studied because of this ---ameliorating effect.

1.6 Fetal Haemoglobin & Hereditary Persistence of HbF

Globin switching is subjected to strict regulation at various levels. The switch from HbF to HbA is normally completed during infancy by approximately 6 months of age. Fetal haemoglobin binds oxygen more strongly than adult haemoglobin which facilitates the oxygen transfer from mother to foetus. The air to blood transfer in the lungs is more efficient than the blood to blood transfer in the placenta and the high oxygen affinity of fetal haemoglobin is no longer required after birth (Miller et al 2005).

Occasionally, in the context of certain pathological conditions or rare mutations, the level of HbF can be elevated. While increased levels of HbF in healthy individuals are of no consequences,increased HbF levels in individuals with SCD and β-thalassaemia are highly beneficial. This ameliorating effect of fetal haemoglobin in β –globinopathies makes the variability and persistence of fetal haemoglobin in adulthood a clinically important feature (Rochette, Craig & Thein, 1994; Weatherall & Clegg, 2001).
The switch from fetal to adult haemoglobin is not complete, in thatadults continue to express low levels of HbF, that are concentrated in a small percentage of erythrocytes referred to as F cells(Boyer et al 1975). Although present at a low concentration, the levels of HbF vary considerably with the majority of individuals having HbF levels of less than 1%. Individuals with higher levels are said to have a persistent increased production of HbF referred to as hereditary persistence of fetal haemoglobin (HPFH). There are two types of HPFH. Pancellular HPFH refers to the even distribution of HbF in erythrocytes, with clearly recognisable increase of 10-35% HbF. These are caused by mutations (deletions or point mutations in the γ-globin promoter) affecting the β-globin gene cluster. Heterocellular HPFH reflects the uneven distribution of HbF among the red blood cells. Individuals with heterocellular HPFH represent the upper tail of the natural continuous distribution of HbF with HbF levels >0.8%. Heterocellular HPFH is

inherited as a complex trait (Forget et al 1998 , Wood et al 2001; Thein & Menzel, 2009).

1.6.1 Stress Erythropoiesis and Reactivation of HbF

At times when the demand for oxygen is greater, such as anaemia and tissue hypoxia (during fetal development), erythroid progenitor cells, induced by an increase in Epo, have the ability to rapidly proliferate, releasing large numbers of erythrocytes into circulation. This process is called ‘stress erythropoiesis. A γ-globin to β-globin switch also occurs within early progenitor cells (Papayannopoulou et al 1986). While we see a small number of F-cells in normal erythropoiesis, during stress erythropoiesis more cells terminate maturation prematurely while still expressing γ-globin, and are released into circulation while still expressing predominantly HbF as F-cells, thus increasing HbF levels (Stamatoyannopoulos et al 2005).

The elucidation of the molecular basis for γ-globin silencing in the adult stage in particular has been the focus of intense investigation, since it has been observed that coinheritance of genetic conditions that confer elevated γ-globin synthesis can significantly alleviate the symptoms of β-haemoglobin disorders.

After the cloning and mapping of globin genes it was reported that γ-globin was DNA methylated in adult erythroid progenitors but it was not methylated in fetal erythroid progenitors (Fritsch et al 1980, Flavell et al 1980). This report led to the clinical trials of 5-Azacytidine that can reverse DNA methylation increasing expression of γ-globin in phlebotomized primates (De Simone et al 1982).

While 5-azacytidine reverses DNA methylation activity, it also has an inhibitory activity on cell cycle; especially in the S-phase. This led to trials using other S-phase inhibitors

such as hydroxyurea, another S-phase inhibitor with encouraging results. Hydroxyurea has been used to treat SCD patients since the early 1990s, and is the only licensed drug for the disease; but seems to have variable effects in increasing HbF in β-thalassaemia patients. Currently many drugs are under investigation for use as HbF inducers but none are close to being used as therapeutics for haemoglobinopathies. This makes understanding the γ-globin to β-globin switch and control of HbF production even more important. Understanding the mechanisms of HbF regulation and control can provide insights on reactivation of HbF for targeted therapies of the β-haemoglobinopathies.

1.7 Quantative Trait Loci Influencing HbF Levels

HbF behaves as a quantitative trait with a continuous spectrum of variation contributed by several factors (Thein & Craig, 1998).Multiple factors contribute to the variance in HbF levels in the population, including age, sex (2%) and environmental factors (9%), but an overwhelming 89% is attributed to genetic factors referred to as quantitative trait loci (QTL) (Garner et al 2000). Cis-acting factors in the β-globin locus account for 10-30% of the trait variation but transacting regulatory factors from other loci play an even more substantial role accounting for more than 50% of the variance.

Previous studies by our group showed that three loci – the *HBS1L-MYB* intergenic region (*HMIP*) on chromosome 6q23, *BCL11A* on chromosome 2p16 and the β globin cluster on chromosome 11 accounts for up to 50 % of the variation in HbF levels in nonanemic Europeans, 35% of the genetic variance in F cell levels is controlled by the two loci; *HMIP* (19%)and *BCL11A* (15%)(Menzel et al2007).
Association of these 3 loci with HbF production has been validated by different groups in different populations: healthy and with β thalassaemia or sickle cell anaemia (Lettre

et al 2008, So et al 2008, Sedgewick et al 2008,Uda et al 2008, Creary et al 2009, Makani et al 2011).

1.7.1 Chromosome 11 Xmn1-Gγ Polymorphism

The *Xmn1*-G*γ* C-T polymorphism occurs at position -158 upstream of the Gγ-globin gene on chromosome 11p. *Xmn1*-G*γ* has long been implicated as a QTL influencing HbF production through clinical and family studies. The T allele was shown to be associated with elevated HbF levels. This polymorphism is present in about 30% of the population, and results in a slight increase in HbF levels in healthy individuals. However, the presence of this polymorphism in SCD and β-thalassaemia patients results in even higher HbF levels (Thein et al1987).

1.7.2 2p15 BCL11A Locus

To identify additional QTLs, our group conducted a genome wide association study (GWAS) in a Northern European cohort of 179 individuals with contrasting extreme F cell values (Menzel et al 2007).

This study picked up the already known HbF associated loci at the *Xmn1-HBG2* site and *HMIP* on chromosome 6q23, and in addition identified a third, previously unknown QTL on chromosome 2p15, close to a gene encoding the zinc finger protein *BCL11A.* (The strongest association lay at a 14 kb region in intron 2 of *BCL11A*, with the second cluster spanning 67 kb in the 3' region of *BCL11A*, 8-7 kb downstream of exon 5 (Menzel et al 2007)(Fig1.11).

The 2p15 QTL was estimated to contribute 15% of overall trait variance for HbF levels (Menzel et al 2007). The peak of association within *BCL11A* was later confirmed by other groups (Lettre et al 2008, Sedgewick et al 2008, and Udaet al 2008).

Studies showed that variants in intron 2 of *BCL11A* that are strongly associated with HbF was significantly more frequent in patients with β-thalassaemia intermedia, leading to the conclusion that these variants influences a milder phenotype. Thus, BCL11A has important effects in haemoglobinopathies as well as non-anaemic individuals (Uda et al*)* .These studies collectively confirm that the genetic variants at the 2p15 QTL influence HbF levels in SCD and lie within intron 2 of the *BCL11A* gene (Figure 1.11).

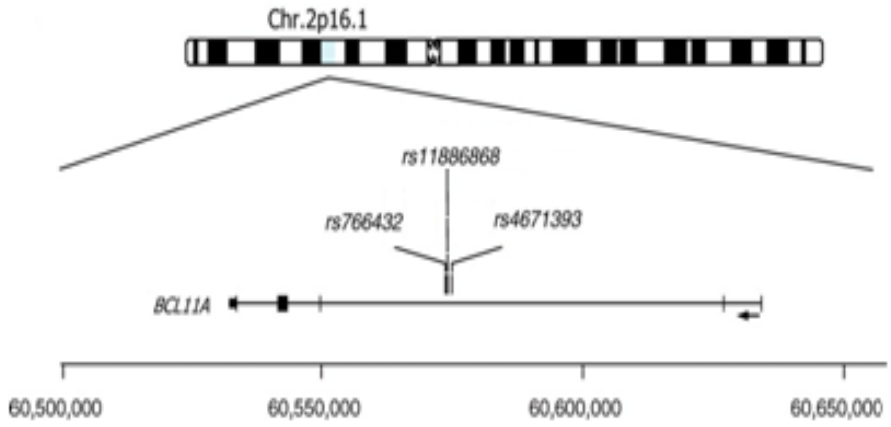


Figure 1.11: QTL in chromosome 2. Highest scoring SNPs indicated at intron 2.

1.7.3 Chromosome 6q QTL

The QTL on chromosome 6q23 was initially identified by linkage analysis in an extended Asian family with β-thalassaemia. The family was identified through the proband that had very mild thalassaemia intermedia despite being homozygous for β^0 thalassaemia with a complete absence of HbA (Thein et al 1989). He was able to compensate for the lack of HbA due to the ability to produce HbF. His steady state total haemoglobin ranged from 11-12 gr/ dl, almost all of which was HbF. Family studies showed that several members of the extended family, with and without thalassaemia, had variable increases in HbF levels of up to 3.4% in healthy members and up to as much as 24% in β-thalassaemia heterozygotes (Thein & Weatherall, 1989, Thein et al 1994).

Segregation studies showed that the genetic determinant for this innate ability to produce HbF was inherited independently of the β-globin gene complex. Family and linkage analysis studies mapped the HbF locus to a region spanning 1.5 Mb on chromosome 6q23. Five protein coding genes, *ALDH8A1*, *HBS1L*, *MYB*, *AHI1*, and *PDE7B* reside in this 1.5-megabase (Mb) candidate interval of 6q23.
None of these genes harboured any causative mutationsin the coding sequence of these genes. (Thein et al 1994; Close et al , 2004).

Further experiments focused on the differences in the 6q23 transcriptome existing between erythroid precursors of individuals with and without high HbF levels.
An expression profile of these five genes in erythroid progenitor cells of 12 high HbF and 14 low HbF individuals showed that *MYB* and *HBS1L* are quantitative trait genes, with variable expression in erythroid progenitor cells of healthy adults. There was a clear correlation of elevated HbF with low levels of MYB and HBS1L in erythroid cells. Furthermore, over expression of *MYB* inhibited γ-globin gene expression when

compared with the parental cell line. However, overexpression of *HBS1L* had no effect on γ-globin gene expression**.**

The 6q HbF QTL was then validated in a large panel of European twins. Higher resolution genotyping in Europeans refined the locus to a region of 79 kb between the *HBS1L* and *MYB* gene. The analysis revealed that this was a major QTL which accounted for 35% of total F-cell variance in the family.

The 6q QTL exists as a set of single nucleotide polymorphisms (SNPs) spanning a region of ≈ 79kb that includes the intergenic region 5^{i} to *MYB and HBS1L , and* 5^{i} part of the *HBS1L gene.*The SNPs are distributed in three linkage disequilibrium (LD) blocks referred to as *HBS1L –MYB* intergenic polymorphism *(HMIP)* blocks 1, 2, and 3. The 3 blocks completely account for the HbF variance attributed to the 6q QTL, with SNPs in *HMIP-2* (24kb) showing the strongest effect and accounting for the majority of the variance (Thein et al 2007).

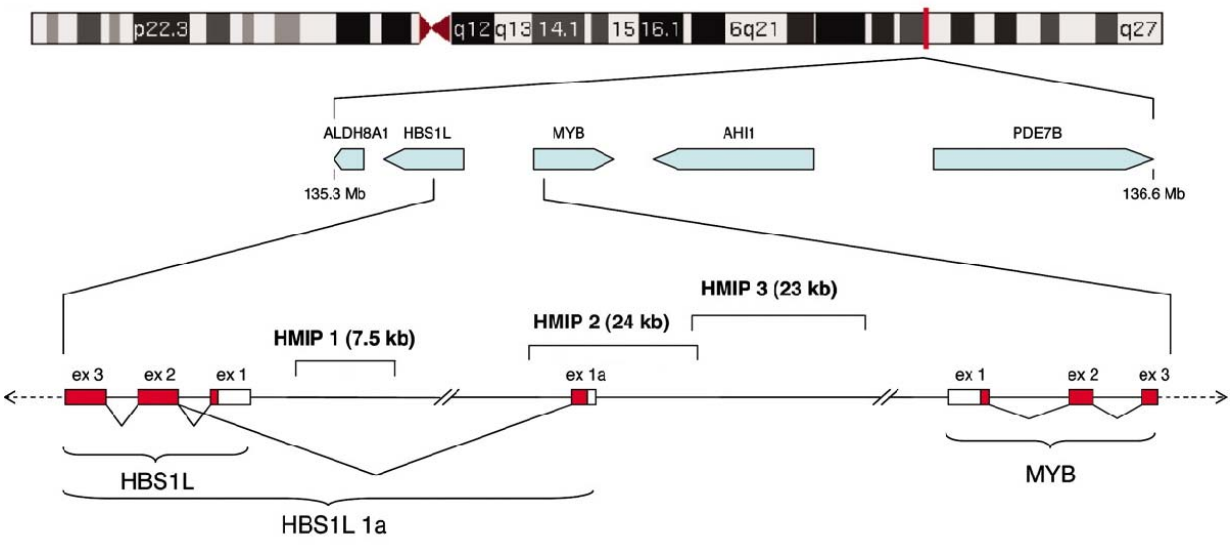


Figure 1.12 : QTL in chromosome 6q23 and *HMIP* blocks 1-3.
Image adapted from Thein et al 2007.

Genetic variants in the *HBS1L-MYB* interval on chromosome6q have been shown to be highly associated, not only with HbFlevels, but also with the control of other erythroid and non-erythroid hematologic traits. Our group showed significant impact of *HMIP* variability on several types of peripheral blood cells: erythrocyte, platelet, and monocyte counts as well as erythrocyte volume and hemoglobin content in healthy individuals of European ancestry (Menzelet al 2007).

Recent work by our group indicated the presence of regulatory elements in *HBS1L-MYB* intergenic region (*HMIP*) evidenced by GATA-1 binding coinciding with strong histone acetylation, RNA polymerase II activity, and erythroid specific DNase I hypersensitive sites. Wahlberg et al *(*2009) identified DNase I hypersensitive sites in *HMIP* block 2 suggesting the presence of regulatory elements in the *HBS1L-MYB* intergenic region further supported by the presence of erythroid specific transcripts. Recent genome-wide datasets generated by the ENCODE consortium validated the presence of DNase hypersensitive sites in erythroid K562 cells and revealed numerous enhancer chromatin signatures. A strong positive correlation between *MYB* expression levels and intergenic regulatory activity could be demonstrated; while no clear correlation with *HBS1L* expression was observed.

Analysis of the region for transcription factor binding and chromatin modification using chromatin immunoprecipitation on microarray (ChIP-chip) indicated five GATA-1 binding sites in block 2 (Wahlberg et al 2009, Jawaid et al 2010).

The mechanism of how the QTLs affect HbF levels is still to be clarified. It would seem that BCL11A has a direct repressive effect on γ-globin gene while that of the *HBS1L-MYB* intergenic region has an effect through regulation of its flanking genes: *HBS1L* and/or *MYB*.

1.8 Aims & Objectives

The overall aim of this project is to carry out functional analysis of the 6q QTL; to investigate and identify the causative SNPs in the region and their mechanism of action. The analysis was concentrated on the *HMIP-2*region which showed the strongest association to HbF levels.

My immediate objectives were:

1. To investigate the best culture method for obtaining homogenous and synchronized cells (Chapter 3).
2. To perform expression profiling of critical transcription factors of interest(Chapter 3).
3. To investigate transcription factor binding at the *HMIP* locus using prediction tools in order to find out if any of the high scoring SNPs alter binding of key transcription factors(Chapter 4).
4. To investigate the potential differential binding of transcription factors to newly revealed SNPs using EMSAs (Chapter 5).
5. To perform ChIP-qPCR experiments to quantify GATA-1 binding to previously demonstrated targets (Chapter 6).
6. To perform ChIP-qPCR experiments to investigate MYB and KLF1 binding to targets suggested by *in-silico* tools.
7. To investigate differential transcription binding with individuals that are +/+ i.e. (homozygote for the presence of SNPs associated with elevated HbF levels) and -/- (homozygote for the absence of SNPs associated with elevated HbF levels) by performing ChIP-qPCR and Allele specific ChIP (Chapter 6).
8. To investigate if there is unequal expression of flanking genes i.e. (MYB and *HBS1L)* in individuals heterozygous for the *HMIP* variants in order to show the direct correlation between the intergenic variants and the expression of flanking genes (Chapter 7).

2 MATERIAL & METHODS

2.1 Erythroid Cell Cultures

Three different methods of liquid culture were performed during the project for primary human erythroid progenitor cells, two of which were adopted from the original method described by Fibach et al (Fibach et al 1989). The other method was adopted from Van Den Akker et al (Haemotologica 2010). Table 2.1 compares key points of three culture methods.

Key Points	**2-Phase Fibach Culture**	**2-Phase Stem-span Culture**	**3-Phase Culture**
Plastic	Cells are incubated in flasks.	Cells are incubated in flasks.	Cells are incubated in petri dishes or 6 well plates for better aeration.
Base Medium	Biological Foetal Calf Serum (FCS) is used as base medium. Batch variable	Semi-synthetic Stemspan is used as base medium. Batch variable	Stemspan is used as base medium. Batch variable
Cytokines	Conditioned medium includes key cytokines needed for proliferation of early progenitor cells.	IL3 replaces conditioned medium for proliferation of early progenitor cells.	Additional cytokines (lipids, IGF1) are included in phase I and phase II to induce proliferation
Media volume	10-25 ml depending on cell counts	10-25 ml depending on cell counts	2-10 ml depending on cell counts
Concentration of cells	Cell concentration is 1million cells /ml during phase I	Cell concentration is 1million cells /ml during phase I	Cell concentration is 5-10million cells /ml during Phase I

			leading to induced cell interactions
T-lymphocyte depletion	Cyclosporin is used to kill T-lymphocytes	Cyclosporin is used to kill T-lymphocytes	Cyclosporin is not used to kill T-lymphocytes. Histopaque layering can be performed to deplete T-lymphocytes
Macrophage depletion	Macrophages are depleted during phase 1 and 2 by changing the flasks	Macrophages are depleted during phase 1 and 2 by changing the flasks	Macrophages are depleted during phase 1 and 2 by changing the petri dishes

Table 2-1: Comparison of three erythroid culture methods

2.1.1 Two-Phase Fibach Culture

50 mls of peripheral blood in EDTA was diluted 1:1 with PBS. This was layered onto a density gradient centrifugation liquid- Histopaque-Ficoll (Sigma-Aldrich, UK) and centrifuged to separate the blood components. The buffy coat containing lymphocytes and progenitor cells was collected and washed with PBS. The washes were performed till the supernatant is transparent.

The cells were counted and cultured in flasks at 1-2 million/ml in Phase I medium [DMEM-Alpha modified medium (Sigma, Poole, UK), 5 U/ml penicillin, 5 µg/ml streptomycin, 1µg/ml cyclosporin A, 2mM L-Glutamine, 10% foetal calf serum (Sigma, Poole, UK) and 10% conditioned medium from cultures of 5637 bladder carcinoma cell line (ATCC)] at 37° in 5% CO_2 for 6 days.

24 hours later the entire culture suspension was transferred to a fresh flask in order to remove adherent monocytes in the first flask. After 6 days in phase I culture, the cells were centrifuged and resuspended in Phase II medium [DMEM alpha modified Medium, 5 U/ml penicillin, 5 µg/ml streptomycin, 1 U/ml human recombinant erythropoietin [EPO; Eprex UK] and 10 ng/ml SCF, 2mM L-Glutamine 30% Foetal calf serum , 10^{-5}M β-Mercaptoethanol, 10^{-6} mol/l dexamethasone ,0.3mg/ml Holotransferrin(ICN) , 1% Bovine serum albumin (Sigma, Poole, UK). The cells were maintained at a concentration of 1 million/ml for up to 15 days.

2.1.1.1 Preparation of conditioned medium

5637 cells were cultured in DMEM alpha modified medium,5 U/ml penicillin, 5 µg/ml streptomycin, 2mM L-Glutamine and 15% Foetal calf serum. Medium was changed after formation of a monolayer and cells were incubated for another week. Supernatant was centrifuged at 1000rpm for 5 minutes, filter sterilised and stored at -20˚C for future use.

This conditioned medium includes cytokines which favours cell growth. It was replaced by stem cell factor in Phase II.

2.1.1.2 Preparation of Bovine Serum Albumin (BSA)

50 gm of BSA (Fraction V) powder was dissolved in deionised water overnight. The dissolved BSA was incubated with 5gm of resin (Bio-Rad) for 30-60 minutes at room temperature. The solution was filtered and incubated with more resin until no colour change of resin was observed which suggest all the chemicals in the solution are taken away with the resin. After the final filtering an equal volume of double concentrated alpha medium (Sigma) was added to the solution and it was stored in 50ml aliquots at -20˚C for future use. BSA was used in the second phase and improper preparation leads to problems such as limited cell proliferation.

2.1.2 Two-Phase erythroid culture using Stem Span

This liquid culture technique was adapted from the original method described by Eitan Fibach (Fibach et al 1989), with modifications as detailed in Jiang et al 2006.
50mls of peripheral blood was collected and centrifuged to separate the blood components. The yellow plasma interface is the component containing lymphocytes and progenitor cells and is called the buffy coat. This was collected and layered onto a density gradient centrifugation liquid- Histopaque-Ficoll (Sigma-Aldrich, UK) and centrifuged to separate red and white blood cells. The lymphocytes and other mononuclear cells remain at the plasma-histopaque interface from where they are collected and washed in PBS. Cells are counted and cultured in flasks at 1-2 million/ml in Phase I media (Stem Span Medium [Stem Cell Technologies, Canada], 10U/ml penicillin, 10µg/ml streptomycin, 1µg/ml cyclosporin A, 25ng/ml interleukin 3 [IL-3; Sigma, UK] and 50ng/ml human stem cell factor [SCF; Sigma, UK]) at 37° in 5% CO_2.
24 hours later the entire culture suspension is transferred to a fresh flask, leaving behind the adherent monocytes in the first flask.

After 7 days in phase I, the culture is moved to a 50ml falcon tube, centrifuged and the cell pellet re-suspended in Phase II medium (Stem Span Medium, 10U/ml penicillin, 10µg/ml streptomycin, 2U/ml human recombinant erythropoietin [EPO; Sigma, Poole UK] and 50ng/ml SCF) and maintained at a concentration of 1 million/ml for up to 20 days.

2.1.3 Three-Phase erythroid culture

This method was described by Van Den Akker et al (Haemotologica 2010) and kindly provided to our group. 50mls of peripheral blood in EDTA was diluted 1:1 with PBS and layered onto a liquid density gradient (Histopaque 1.077) (Sigma-Aldrich, UK) and

centrifuged to separate the blood components. The buffy coat containing lymphocytes and progenitor cells was collected and washed with PBS.

Cells were counted and cultured in 6 petri dishes at 5-10 million/ml in Phase I StemSpan medium containing: IL-3 – 1ng/ml, human recombinant erythropoietin [EPO; Eprex UK] – 2U/ml, Stem cell factor – 40ng/ml, Dexamethasone - 1μM, Insulin growth factor 1(IGF1) – 40ng/ml, Cholesterol rich mix 50μg/ml, 5 U/ml penicillin, 5 μg/ml streptomycin for 5 days. Daily petri dish changes were performed to minimize number of monocytes that adhere to the petri dishes. On day 2 cells were centrifuged and resuspended in fresh medium.

After 6 days in phase I, the culture is moved to a 50ml falcon tube, centrifuged and the cell pellet re-suspended in Phase II media. Cells are counted and cultured in petri dishes at 0.5-1.0 million/ml in Phase II StemSpan media containing: human recombinant erythropoietin [EPO; Eprex UK] – 2U/ml, Stem cell factor – 40ng/ml, Dexamethasone - 1μM, Insulin growth factor 1(IGF1) – 40ng/ml, Cholesterol rich mix 50μg/ml, 5 U/ml penicillin, 5 μg/ml streptomycin for 7-25 days.

Differentiation can be induced after a minimum of 7 days in phase II which involved collecting the cells and culturing them in phase III StemSpan medium containing: Human recombinant erythropoietin [EPO; Eprex UK] – 10U/ml, Insulin growth factor 1(IGF1) –, 5 U/ml penicillin, 5 μg/ml streptomycin, Holotransferrin 1mg/ml ,Insulin 10ng/ml,thyroid hormone 1μM , 3 %human OAB serum (male) for 2-7 days. Phase III proved to be very crucial; since the very large amount of erythropoietin in the medium leads to accelerated differentiation of the erythroid cells which makes it really hard to harvest them at specific stages. We needed cells at earlier stages i.e. (proerythroblasts, basophilic normoblasts). Because of this cells were not transferred to phase III for induced differentiation.

2.1.4 Cell morphology

Cytospins of erythroid cells were made by spinning $1x10^5$ cells in a 100-200µl volume onto glass slides. Slides were then stained using Giemsa staining set (Hema "Gurr", VWR, Lutterworth, UK.), and viewed with 40x objective under the microscope (Zeiss Axiolab).

2.1.5 Flow Cytometry

Fluorescence activated cell sorting (FACS) - a specialized type of flow cytometry was used to test erythroid cell purity using cell surface markers GPA and CD71. Glycophorin A (GPA) and CD71 are human red blood cell membrane proteins (Furthmayr et al 1975). GPA protein is highly expressed on the surface of mature erythroid cells, while CD71 is expressed in proliferating eryhtroid cells. Detection of GPA and CD71 indicates the purity of late erythroid progenitor cells in the culture. 2 µl Fluorescent antibodies against human CD71 (fluorescein isothiocyanate [FITC] conjugated, BD Biosciences, Oxford, UK) and anti-human Glycophorin A (GPA, R-phycoerythrin [RPE] conjugated, DAKO, Glostrup, Denmark) were added to $1x10^6$ erythroid cells washed in PBS and resuspended in 100µl PBS. Both antibodies were added for a double staining and one sample is without antibody to serve as a negative control. Cells and antibody were incubated for 30 minutes in the dark, after which they were washed twice with PBS and resuspended in a final volume of 1ml PBS. FACS analysis of 10,000 cells was acquired using Becton Dickinson flow cytometer and Cell Quest software (Becton Dickinson, Oxford, UK).

2.2 K562 Cell Culture

K562 cell lines were maintained in RPMI-1620 medium (Sigma-Aldrich, Poole, UK) with the addition of 10% of fetal calf serum (FCS) (PAA-laboratories, Somerset, UK]), 2 mM of L-glutamine (Sigma-Aldrich, Poole, UK), 0.1 mg/ml of streptomycin and 18 units/ml

of penicillin (Sigma-Aldrich, Poole, UK). Cells were cultured in incubators at 37 °C and 5 % CO_2 and kept at a concentration between 0.5-1x10^6 cells/ml.

2.3Western blotting

2.3.1 Protein extraction

5x10^6 cells were collected, centrifuged and re-suspended in 250µl RIPA cell lysis buffer (20mM Tris-HCl, 150mM NaCl, 1% NP-40, 0.5% sodium deoxycholate, 1mM EDTA 0.1% SDS) and 12.5µl 20X protease inhibitor (1 Complete EDTA-free Protease Inhibitor Cocktail tablet [Roche, UK] dissolved in 500µl H_2O).

2.3.2 Nuclear Protein Extraction

Proteins were extracted using the Thermoscientific NE-PER Nuclear and Cytoplasmic Extraction Kit**.** Before starting the experiment protease inhibitors were added to the reagents in order to prevent any possible degradation of proteins. Cells in the suspension were harvested by centrifuging at 500xg for 5 minutes. The cell pellet was washed with PBS.1-10 x 10^6 cells were transferred to a 1.5 ml microcentrifuge. The cells were pelleted by centrifugating at 500xg for 3 minutes. Ice cold cytoplasmic extraction reagent I (CER I) was added to the cell pellet. The ratios in table 2.2 were applied.

Packed Cell Volume (µl)	CER I (µl)	CER II (µl)	NER (µl)
10	100	5.5	50

Table 2-2: Optimal volumes of reagents.

Reagent volumes were determined according to the packed cell volume. Each reagent was added according to the ratios shown in the table.

The cell pellet was suspended by vigorous vortexing for 15 seconds followed by 10 minute incubation on ice. Ice cold CERII was added and the tube was vortexed vigorously for 5 seconds followed by 1 minute incubation on ice. The tube was vortexed for another 5 seconds and the cells were centrifuged at 16000xg for 5minutes. Supernatant containing the cytoplasmic extract was discarded and the pellet re-suspended in ice cold nuclear extraction reagent (NER) buffer according to previously mentioned ratios. The cells were incubated on ice for 40 minutes and vortexed vigorously every 10 minutes for better suspension. After pelleting the cells by centrifugation at 16000xg for 10minutes the supernatant containing the nuclear proteins were transferred to a pre-chilled tube, snap frozen by liquid nitrogen and stored at -80˚C after quantification.

2.3.3 Quantification of proteins

Proteins in nuclear extracts were quantified using a NanoDrop Spectrophotometer ND-1000 (NanoDrop Technologies, Wilmington, USA). Nanodrop program AB280 was calibrated with controls and quantification was performed.

2.3.4 Protein electrophoresis and transfer

Samples were denatured for 10 mins in a heat block at 95°C. 25µg of each protein (25µl) and 10µl SeeBlue protein marker (Invitrogen) were loaded on a NuPAGE 10% Bis-Tris precast gel (Novex, Invitrogen) in MOPS SDS running buffer (Novex, Invitrogen). The gel was run overnight at 20V using the Novex Invitrogen Surelock X-Cell electrophoresis system, to separate proteins. Proteins were transferred from gel to nitrocellulose membrane (Amersham Hybond ECL membrane, GE Healthcare) for 2 hours at 55V (X-Cell blot module, Novex Invitrogen) in transfer buffer (3.04g Tris, 14.5g glycine, 200mls methanol 800mls H_20).

The nitrocellulose membrane was transferred to a 50ml falcon tube and non-specific binding was blocked with a buffer containing 5% dry milk (Marvel) in PBS-Tween (5 PBS tablets dissolved in 5mls Tween-20 and 995mls H_20) for a minimum of 2 hours at 4°C with constant rolling.

2.3.5 Antibody probing

The membranes were then incubated in 50ml falcon tubes overnight in a buffer containing 5% milk and the primary antibody at 4°C with constant rolling.
The following morning after washing 3 times for 10 mins in PBS-Tween the membranes were incubated in a buffer containing 5% milk and the secondary antibody at room temperature for 1 hour 15mins with constant rolling. This was followed again by three 10 min washes in PBS-Tween.

2.3.6 Target protein detection

Target proteins were detected by an enhanced chemiluminescence kit (Amersham ECL Plus western blotting detection system, GE Healthcare). Following the manufacturer's instructions, the membrane was incubated with ECL reagents for 5 mins and drained off. The membrane was wrapped in Saran wrap and placed in an exposure cassette. In a dark room the membrane was exposed to high performance chemiluminescence film (Amersham Hyperfilm ECL, GE Healthcare) for varying amounts of time (between 10 seconds to 3 mins) and the film was developed using the Compact X4 X-Ray film processor (Xograph Imaging Systems, UK)
To re-probe with another primary antibody, membranes were re-incubated in a buffer containing 5% milk to block primary antibody before incubating in another antibody overnight as before.

2.4 Transcription Factor Predictions at Chromosome 6q QTL

In-silico prediction for transcription factor binding was performed on the SNPs in the chromosome 6q QTL. Transcription factor prediction tools (matinspector and jaspar) were used to predict transcription factor binding sites around the SNPs. Further comparison of predicted binding sites with the alternative alleles of candidate SNPs was performed to investigate if the SNPs alter binding of any transcription factors.

2.4.1 Matinspector

Each SNP is uploaded to the tool with 25bp of flanking sequences from each side as shown in figure 2.1.

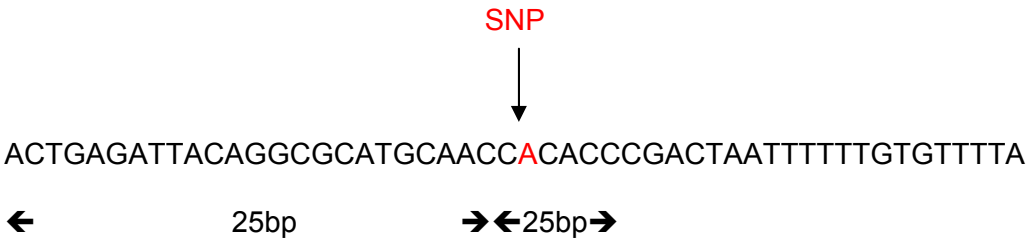


Figure 2.1: Sequence uploaded to prediction tools encompasses the SNP flanked by 25 base pair of nucleotides at each side.

Matrix similarity threshold was set high i.e. (0.700) to make the results easier to interpret since setting the threshold low will lead to excessive amount of matches. Predictions were made for all transcription factor families in the matrix (TRANSFAC 6.0).The same protocol was repeated with the wild type sequence. The results were compared to see if any transcription factor binds differentially to the sequences in the presence of alternative alleles of the SNPs. Any major difference affecting matrix similarity was presented as demolishment of binding site of the transcription factor suggesting differential binding.

2.4.2 Jaspar (Consite)

Transcription factor predictions highlighted some SNPs which altered the binding of transcription factors. Further predictions were performed for those SNPs to confirm altered binding. Consite, a prediction tool belonging to Jaspar (another transcription factor database) was used.

Each SNP is uploaded to the tool with 25bp of flanking sequences from each side as shown in figure 2.1. Matrices belonging to the transcription family of interest were chosen. Some transcription factor families were not present in the matrices available.

2.5 Electrophoretic Mobility Shift Assays (EMSA)

2.5.1 Oligo Design

25 base pair oligonucleotides were designed where SNPs and predicted binding motifs are at the center of the oligo to provide appropriate amount of flanking sequence for efficient transcription factor binding (Appendix).

Table 2.1 in appendix two shows the designed oligonucleotides and SNPs covered by the oligonucleotides. Oligonucleotides were either unlabeled or biotin labeled.

2.5.2 Annealing of oligonucleotides

Oligonucleotides were suspended in annealing buffer (10mM Tris, 50mM NaCl, 1mM EDTA ph7.7) and equal molar concentrations of forward and reverse oligonucleotides were added to a 1.5 ml tube to make a 100 pmoles/µl stock. The mix of forward and reverse oligonucleotides was incubated at 95^{o}C in a heat block for 3 minutes and cooled down to room temperature slowly on a bench top.

2.5.3 Preparation of Binding reactions

Initial binding reactions were prepared with controls provided with the Thermo Scientific Lightshift Chemiluminescent EMSA Kit (Pierce). The kit included nuclear extracts from Epstein - Barr Virus Nuclear Antigen expressing (EBNA) cell lines where the control protein was highly expressed. Unlabelled and biotin labelled nucleotides were also provided. Binding reactions were prepared according to manufacturer's instructions (Table 2.3). After optimization, binding reactions were prepared with nuclear extracts from K562 cells or human primary erythroid cells, and oligonucleotides that encompass the SNPs of interest.

Binding reactions for EMSA experiments				
Component	**Final Amount**	Binding Reactions		
		No extract	Binding reaction	Cold competition
Ultrapure Water	----	12 µl	11 µl	9 µl
10X Binding Buffer	1X	2 µl	2 µl	2 µl
50% Glycerol	2.5%	1 µl	1 µl	1 µl
100 mM MgCl2	5 mM	1 µl	1 µl	1 µl
1 µg/µl Poly (dI•dC)	50 ng/µl	1 µl	1 µl	1 µl
1% NP-40	0.05%	1 µl	1 µl	1 µl
Unlabeled oligonucleotide	4 pmol	-----	-----	2 µl
Nuclear Extract	1 Unit	-----	1 µl	1 µl
Biotin labeled oligonucleotide	20 fmol	2 µl	2 µl	2 µl
Total Volume	----	20 µl	20 µl	20 µl

Table2-3: Binding Reactions for EMSA

Three different binding reactions were prepared. The first reaction (negative control) was prepared without any extract. The second reaction was prepared with the extract provided with the kit. In the third reaction, unlabelled oligonucleotides were included as cold competitors to differentiate between specific and nonspecific protein-nucleic acid complexes. Reactions without cold competitors were incubated at room temperature for 20 minutes. Cold competitions were performed by addition of 20-200 folds unlabelled oligo as the cold competitor and the reaction was incubated 5 minutes at room temperature followed by 20 minute incubation at room temperature with the biotin labelled oligo. The details for each binding reaction will be presented at chapter 5.

2.5.4 Gel Preparation and Pre-running

Native polyacrylamide gel was prepared with the following recipe : 2ml 5X TBE , 4ml %30 bis-polyacrylamide , 0.625ml %80 glycogen ,13.375 ml water, 0.300mlAPS , 20ul TEMED. Gel was cast and the electrophoresis unit filled with 0.5X TBE. Pre-run was performed at 100 V for 30-60 minutes using NOVEX-Invitrogen, X-Cell Sure Lock, Invitrogen).

2.5.5 Electrophoresis of Binding Reactions

After the incubation, 5µl of 5X loading dye (thermoscientific) was added to each 20µl reaction. 20µl of each sample was loaded to the wells and electrophoresis reaction was performed at 100V for 90-120 minutes at room temperature.

2.5.6 Electrophoretic Transfer of Binding Reactions

Sponges, biodyne membrane and filter papers were soaked in 0.5X TBE before the transfer. Transfer sandwich was prepared as shown in figure 2.4. Transfer was performed at constant amperage of 380mA for 60 minutes at 4˚C.

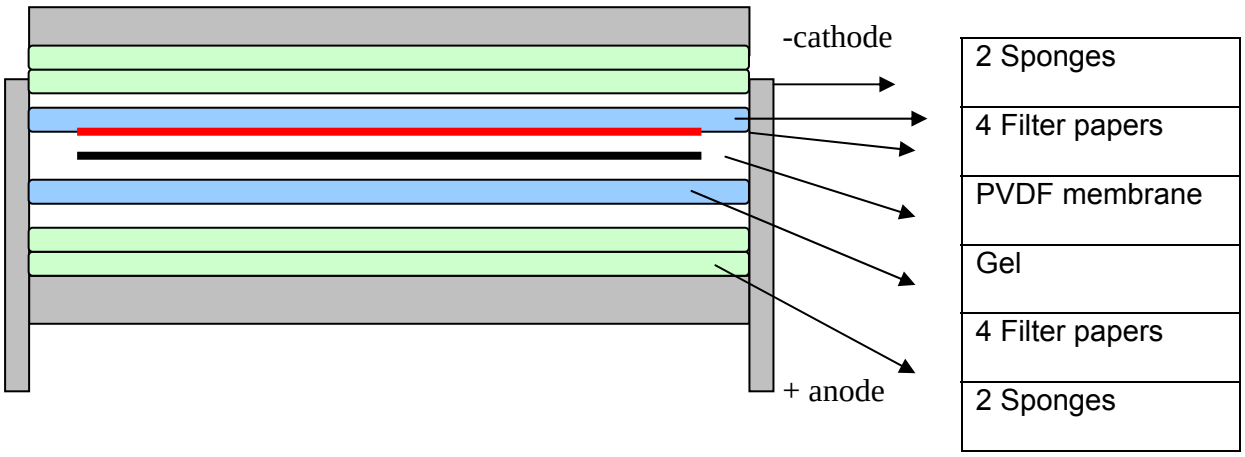


Figure2.2: Set up of transfer chamber sandwich for the transfer of protein-DNA complexes.

2.5.7 Cross-linking of Transferred DNA to Membrane

Cross-linking was performed at 120 mJ/cm^2. Commercial UV-light instrument equipped with 254 nm bulbs was used as the cross-linker.

2.5.8 Detection of Biotin-labelled DNA by Chemiluminescence

Blocking buffer (thermoscientific) and 4X wash buffer (thermoscientific) was warmed to 37˚C before performing the detection. Membrane was blocked by adding 20 ml of Blocking Buffer and incubating for 15 minutes at room temperature with gentle shaking. Conjugate blocking buffer was prepared by adding 66.7 µl Stabilized Streptavidin-Horseradish Peroxidase Conjugate to 20 ml Blocking Buffer (1:300 dilutions). The blocking buffer was replaced with conjugate blocking buffer and membrane was incubated for 15 minutes at room temperature with gentle shaking. 1X was buffer was prepared by adding 30ml 4X wash buffer to 90ml ultrapure water (sigma).Membrane was transferred to a new container and 4 washes were performed using 25ml of 1X wash buffer and incubating for 5 minutes at room temperature with gentle shaking.

Membrane was transferred to a new container and 30ml Substrate Equilibration Buffer was added. Membrane was incubated for 5 minutes at room temperature with gentle shaking.Working Solution was prepared by adding 6 ml Luminol/Enhancer Solution to 6 ml Stable Peroxide Solution. Membrane was removed from previous buffer and excess buffer on the membrane was removed carefully. Membrane was placed DNA side down onto a puddle of the Working Solution and incubated 5 minutes at room temperature.

Membrane was removed and excess buffer was taken away gently by using a paper towel. Membrane was wrapped in Saran wrap and bubbles and excess buffer on the membrane was pushed out of the wrap by rolling.

Exposure was performed under red light. A piece of film (Kodak Biomax light film chemiluminescence for biological imaging) was placed on top of the membrane and saran wrap and the cassette closed firmly. The membrane was left to expose in the cassette for an appropriate time varying from 5 seconds to 2 minutes depending on the strength of the signal.

2.6 Chromatin immunoprecipitation (ChIP)

ChIP experiments were carried out using the EZ-ChIP^{TM} kit (cat# 17-371) according to manufacturer's protocols with minor modifications (modifications kindly provided by Dr. David Garrick from the MRC molecular haematology unit, Oxford). All protease inhibitors used in the protocols were complete mini EDTA free tablets (Sigma-Aldrich, UK).

2.6.1 Cell-protein cross linking

The process of cross-linking links DNA to protein so that the DNA can be co-precipitated with the protein using a specific antibody. $5x10^7$ erythroid cells were taken for cross-linking in 10ml growth medium with 1% formaldehyde (Sigma-Aldrich, UK), and incubated for 10 min at room temperature with rolling. 125mM Glycine was added to quench the formaldehyde and then cells were washed in 5ml of cold PBS with protease inhibitors and centrifuged at 1000 rpm for 10 min at 4°C, twice. The pellets were divided into 5 aliquots of $1x10^7$ cells and stored at -80°C.

2.6.2 Cell lysis

$5x10^7$ cross-linked cells were lysed in 1ml SDS-lysis buffer (EZ-Magna ChIP™ Kit (cat. # 17-409) (Upstate-Millipore, UK) with protease inhibitor and incubated on ice for 15 min, vortexing every 5 mins. The cells were spun at 800 x g for 5 mins at 4°C and the pellet re-suspended in 1ml nucleus lysis buffer (EZ-Magna ChIP™ kit) with protease inhibitors.

2.6.3 Chromatin fragmentation by sonication

Cell/nuclei lysates ($5x10^7$ cells/ml equivalent) were sonicated with a Vibra Cell sonicator (Sonics, US) set at power 2 for 15x10 seconds using a small probe (2mm in diameter) to achieve DNA fragment lengths of 500-1000bp.

The samples were kept on ice during sonication and left to rest on ice for 30secs in-between each round of sonication to avoid overheating with protein denaturation as a consequence. To remove insoluble material, samples were spun at 10,000 x g for 10 min and the supernatant was moved to a new tube.

2.6.3.1 Gel analysis of sonication

For analysis of sonication efficiency, 5µl of sonicated material was collected and de-cross-linked over night at 65 °C after the addition of 90µl of H_2O and 4µl 5M NaCl. Following de-crosslinking, 1µl RNase A (10µg/µl) was added and the sample was incubated at 37 °C for 30-60 min to degrade RNA. Subsequently 2µl of 0.5M EDTA, 4µl Tris-HCl (1 M, pH 6.5) and 1µl Proteinase K (10µg/µl) was added followed by 2 h incubation at 45 °C to degrade proteins. 10µl of chromatin samples were analysed on 1.5 % w/v agarose gels.

2.6.4 Immunoprecipitation (IP)

Sonicated chromatin was diluted 10 times in ChIP dilution buffer supplied in the kit that also provided protease inhibitors. 2 x 10µl of sample was removed and stored at 4°C as input sample and kept until the following day.

10-15µg of antibody was added to the diluted chromatin (equivalent to $1x10^7$ cells) for each immunoprecipitation (IP) and 40µl of protein-G magnetic beads (EZ-Magna ChIP™ kit) were added and incubated overnight at 4°C on rotation. The beads were pelleted in the cold room (to keep proteases cold) using a magnetic tube rack and the supernatant discarded. Beads were then washed by re-suspending the bead pellet in 1ml of wash buffer supplied in the kit, incubated for 5 min at 4°C and pelleted again by standing for 1 min on the magnet, the wash was discarded and pellet re-suspended in the subsequent wash buffer and incubated again. Washes were performed in the following order:

1. Low salt immune complex wash buffer (0.5ml)
2. High salt immune complex wash buffer (0.5ml)
3. LiCl immune complex wash buffer (0.5ml)
4. TE buffer (0.5ml)

The last TE wash was performed at RT. In order to elute the bound chromatin, 200µl of elution buffer (containing 10µg/ml Proteinase K) was added to re-suspend each sample and the two input samples from the previous day. All samples were incubated at 62°C for 3 hours with constant rotation. In this step the elution buffer separates the chromatin from the antibody and beads, the heat reverse-crosslinks the chromatin protein from the DNA and the Proteinase K treatment destroys the proteins. Samples were then incubated at 95°C for 10 min and the beads were pelleted using the magnetic rack, the DNA is now in the supernatant which is transferred to a fresh tube.

2.6.5 Purification and ethanol precipitation of ChIP material

200µl of 1:1 phenol/chloroform-isoamylalcohol mixture was added to the final supernatant that contains the DNA. The samples were vortexed and spun at 13000 rpm for 1 min. The aqueous top layer was transferred to a fresh tube containing 20µl of 3M sodium acetate and 20µg glycogen (Roche, UK). DNA was precipitated at -20°C overnight and pelleted by centrifugation at 14 000 rpm for 40 min at 4°C. Pellets were washed with 150µl 70% EtOH followed by centrifugation at 14 000 rpm for 10 mins at 4°C. Dried pellets were re-suspended in 20µl of water.

2.6.6 Analysis of ChIP Material

Immunoprecipitated chromatin was analysed with different methods depending on the experiment and the analysis include: Real time quantitative PCR (qPCR), Snapshot and Restriction Enzyme Digestion.

2.6.6.1 Analysis by real time PCR

Purified ChIP DNA was measured using the NanoDrop® Spectrophotometer ND-1000(NanoDrop Technologies, USA) and analysed for fold enrichment with control primers and primers for targets (Appendix) using real time SYBR Green PCR which

measures double stranded DNA generated. 0.4-0.5µl (~20ng) of ChIP sample and input sample was mixed with 2-4pmol each of forward and reverse primer, and 5µl of SYBR Green PCR master mix (Applied Biosystems, UK) in a 10 µl reaction volume. All samples were analysed in duplicate using standard real time PCR conditions; 50°C for 2 min, 95°C for 10 min, 95°C for 15 sec and 60°C for 1 min repeated for 40 cycles.

Fold enrichment for an antibody (ab) at a specific primer target (amplicon) in ChIP samples was calculated relative to isotype control (IgG) and normalised against negative control NEFM using the following formula:

Nonspecific adjustment (dCt) = Ct Ab – Ct IgG

Fold Enrichment: 2^{-dCt}

Fold enrichment for a negative control is also calculated (ideally should be 1). Fold enrichment of the target site is then normalised to fold enrichment of the negative control.

2.6.6.2 Analysis by Snapshot

To quantify allelic differences of transcription factor binding around target regions encompassing SNPs of interest and key transcription factors Snapshot analysis of ChIP was performed as described in Chapter 2.8.Snapshot reactions are electrophoresed on a DNA Sequencer (Applied Biosystems, UK). Peak heights of each allele were determined using Gene Marker V2 2.0 demo software (Soft Genetics LLC) and ratio of the heights of allele was calculated. Input samples were analysed by Snapshot in parallel with ChIP samples from the same individual. For each experiment, the ratio of alleles for input and ChIP material were calculated independently. Ratio of the alleles in the ChIP material was then normalized to ratio of the alleles in the input material.

2.6.6.3 Analysis by Restriction Enzyme Digestion

Purified Immunoprecipitated chromatin and input was PCR amplified. PCR products are column purified with Qiagen Nucleotide Removal Kit (Qiagen) according to manufactures instructions.

The purified PCR products are digested with an appropriate restriction enzyme, in this case 2 units of MaeIII (Roche), at 55°C for 3 hours and digested PCR products of ChIP and input were electrophoresed on a 3% w/v agarose gel. The gel was scanned with Image Quant 5.2 and the density of each band was calculated for ChIP and input separately resulting in ratios for digested and undigested bands. The ratio with ChIP material was normalized to the ratio of digested and undigested bands of input.

2.7 Gene expression

All DNA, RNA and protein concentrations were measured on a Nanodrop® 2000 spectrophotometer (Thermo Scientific). The thermal cycler used throughout was the PTC-200 Peltier thermal cycler (MJ Research).

2.7.1 Total RNA extraction and DNase Treatment

Total RNA was extracted from cells using TRIReagent (Sigma, Poole, UK) according to the manufacturer's instructions. One ml of TRI-reagent was used to lyse between 5-10 $x10^6$ cells; for less than $5x10^6$ cells, a minimum of 0.5ml of TRI-reagent was used. RNA was resuspended in 100µl of DEPC-treated water (Applied Biosystems, Ambion) and stored at -80°C. After quantification by Nanodrop the RNA samples were DNase treated using Ambion® TURBO DNA-free™ Kit (Applied Biosystems, Ambion) to remove genomic DNA prior to reverse transcription. 5µl of 10X TURBO DNase buffer and 1µl of TURBO DNase was added to 5µg of RNA in a 50µl reaction volume and incubated at 37°C for 30 mins. 5µl of DNase inactivation reagent was added to the

reaction and incubated for 5 mins at room temperature and mixed at least 3 times over the 5 min incubation period. Finally the DNase I-treated sample was centrifuged at 1000 x g for 1.5 min and transferred to a fresh tube.

2.7.2 Reverse Transcription

First strand cDNA was generated from RNA by using Superscript^{TM} III Reverse Transcriptase (Invitrogen, Paisley, UK). Reactions were set up with 500ng RNA, 1µl sigma random hexamers (0.5µg/µl) and 1µl dNTPs (10mM) in a 13µl volume and incubated on a PCR block at 65˚C for 5 min. 4µl of 1st strand buffer (5x), 1µl DTT (0.1 M), 1µl Superscript^{TM} III RT (200U/µl) and 1µl water were added to a final volume of 20µl. The reactions were incubated on a PCR block at 50˚C for 60 mins followed by 70˚C for 15 mins and cooling to 4˚C.The cDNA was diluted with water to a total volume of 50µl and stored at -20°C.

2.7.3 Analysis by Real Time PCR

Custom designed primers and probes for TaqMan real time PCR analysis were ordered from Applied Biosystems (UK) (Appendix). Primer probe mixes for *BCL11A* (Hs00256254- Invitrogen, UK), *MYB* and housekeeping genes *HPRT1* (Hs99999909) and *GAPDH* (Hs99999905) were ordered as ready-made gene-expression assays from Applied Biosystems.10µl reactions were set up with 2µl cDNA, 1.5pmol each of forward and reverse primer, 2pmol of fluorescent probe and 5µl of TaqMan PCR master mix (Applied Biosystems, UK) . In case of *KLF1* and *GATA-1*, SYBR green real time PCR was performed with custom designed primers (Appendices) in a volume of 10µl that included 2 µl cDNA, 1.5pmol each of forward and reverse primer, and 5µl of SYBR green master mix (Applied Biosystems, UK). All prepared reactions were run on an ABI Prism 7900HT Sequence Detection system (Applied Biosystems, UK) under the following conditions;

Step1) 50°C for 2 min,

Step 2) 95°C for 10 min,

Step 3) 95°C for 0.15 min and 60°C for 1 min repeated for 40 cycles.

Real time PCR measures the number of cycles to reach the threshold of fluorescence detection. The cycle at which the threshold is reached is called the threshold cycle or C_T, which provides a value for the starting copy number of a DNA target. A lower C_T value means fewer amplification cycles were required to reach the threshold of fluorescence detection and therefore the higher the starting amount of our specific DNA sequence. Expression of target gene and the housekeeping gene were each normalized to appropriate negative control. Relative expression of target gene was then calculated relative to the housekeeping gene.

2.8 Snapshot

2.8.1 Standard PCR and Enzyme Purification

DNA was specifically amplified by standard PCR. 25µl reactions were set up with 5µl of DNA, 10µM each of forward and reverse primers (Appendix), 10mM dNTPs, 2.5µl MgCl2, 2.5µl 10X buffer and 0.4µl of AmpliTaq Gold. The reactions were run on a thermal cycler under the following conditions;

Step1) 94°C for 10 min

Step 2) 94°C for 0.30 min

Step 3) 55°C for 0.30 min

Step 4) 72°C for 0.30 min

Step 5) Repeat steps 1-4 for 30 cycles

Step 6) 72°C for 2 min

PCR product was initially visualized by agarose gel electrophoresis of a small amount of the product. Once PCR quality was confirmed, all the remaining PCR product was enzyme purified with 2 units Exonuclease I (New England Biolabs) and 5 units rAPid Alkaline Phosphatase (Roche) at 37 °C for 60min followed by 75 °C for 15min to remove free primers and deoxynucleotide triphosphate (dNTP).

2.8.2 Primer extension

Primer extension reactions was prepared in a total volume of 10 µl with2 µl enzyme purified PCR products, 10µM of appropriate extension primer (Appendix) and 1 µl of Abi Prism Snapshot Mix (Applied Biosystems, UK) according to manufacturer's introductions .Single base extension was performed using conditions: 96 °C for 10sec, 50°C for 5sec, 60°C for 30sec repeated for 25 cycles.

2.8.3 Electrophoresis and data collection

2 µl of primer extension reaction was mixed with 8µl of formamide, vortexed well and spun down. Samples were then transferred to a 96-well plate and pulse spun to 1000rpm. Samples on the plate were then denatured for 2 minutes at 95˚C on the thermal cycler. The plate was then set up and run on the sequencing machine 3130xl Genetic Analyser (ABI Prism by Applied Biosystems) as per the manufacturer's instructions. Data was analysed using Gene Marker V2 2.0 demo software (Soft Genetics LLC).

Ratio of the peak heights of each allele of test samples i.e. (ChIP or cDNA) and controls i.e. (Input or genomic DNA) were determined .Ratio of the alleles (relative expression) of the test samples was then normalized to controls.

2.9 DNA Sequencing

2.9.1 Standard PCR and column purification

DNA was specifically amplified by standard PCR. 25µl reactions were set up with 5µl of DNA, 10µM each of forward and reverse primers (Appendix 1, Table 11.3, 11.4 & 11.7), 10mM dNTP, 2.5µl MgCl2, 2.5µl 10X buffer and 0.4µl of AmpliTaq Gold. The reactions were run on a thermal cycler under the following conditions;

Step1) 94°C for 10 min

Step 2) 94°C for 0.30 min

Step 3) 55°C for 0.30 min

Step 4) 72°C for 0.30 min

Step 5) Repeat steps 1-4 for 30 cycles

Step 6) 72°C for 2 min

PCR product was initially visualized by running a small amount of the product on an agarose gel. Once PCR quality was confirmed, all the remaining PCR product was column purified as per the protocol outlined in the Promega Wizard SV Gel and PCR Clean-up System instruction booklet. DNA concentration was then measured by the Nanodrop®.

2.9.2 Cycle sequencing

Sequencing reactions were carried out using the ABI PRISM Big Dye Terminator v3.0 Ready Reaction Cycle Sequencing Kit following protocols as outlined in the user's manual. 20µl reactions were set up with 200-300ng of DNA, 3.2pmol primer and 4µl of 5 X buffer and 2µl of Terminator Ready Reaction Mix. Reactions were run on a thermal cycler according to manufacturer's protocol (Applied Biosystems, Warrington, UK).

2.9.3 Electrophoresis and data collection

Reactions were transferred to 1.5ml eppendorf tubes containing 16µl deionized water and 64 µl of 95% ethanol. Samples were vortexed and incubated at room temperature for 15 mins, followed by a 20 minute spin at maximum speed. Supernatant was discarded and sample washed with 400µl of 70% ethanol and spun again for 10 mins. Supernatant was again discarded and pellets dried at room temperature.

The pellets were re-suspended in 20µl of formamide, vortexed well and spun down. Samples were then transferred to a 96-well plate and pulse spun to 1000rpm. Samples on the plate were then denatured for 2 minutes at 95˚C on the thermal cycler. The plate was then set up and run on the sequencing machine 3130xl Genetic Analyser (ABI Prism by Applied Biosystems) as per the manufacturer's instructions. Data was analysed using the DNA sequence analysis programme Sequencher® version 4.6 (Gene Codes Corporation, Ann Arbor, MI USA http://www.genecodes.com).

2.10 Allele specific Gene expression

2.10.1 Informative SNP Selection and Genotyping

Informative SNPs were selected according to their frequency i.e. (> 0.300). Ideally exonic SNPs should be selected as informative SNPs but because of the lack of high frequency exonic SNPs in *MYB* gene an intronic SNP rs210796 with frequency >0.300 was selected. In the case of *HBS1L* gene an exonic SNP rs13064 with frequency > 0.300 was present and it was selected as the informative SNP for allele specific expression analysis. All selected SNPs were genotyped as the part of HapMap project.

Healthy unrelated adults of diverse ethnic backgrounds were recruited. HbF levels were measured by high-performance liquid chromatography (BioRad Variant; BioRad, Hemel Hempstead, UK) and F cells as previously described (Jiang et al Blood 2006) using blood in EDTA .Genomic DNA was isolated from peripheral blood of these

individuals and genotyped for the SNPs in block 2 of the HBS1L – MYB interval of 6q, for SNP *rs210796* in intron 4 of the *MYB* gene and the SNP rs13064 in exon 17 of*HBS1L* gene. Individuals were genotyped with Snapshot as described (2.7) and results are validated with sequencing of genomic DNA from the same individual (2.8).

2.10.2 Erythroid Cultures and cDNA preparation

Individuals with the appropriate genotypes for SNPs in *HMIP-2, HBS1L and MYB* were recruited and erythroid cultures were performed using three-phase erythroid culture method as previously described (2.1.3).

Total RNA was extracted from erythroid progenitors at phase II day 4 using TRIReagent (Sigma, Poole, UK) as described in 2.6.1.Total RNA was DNAse treated (2.6.2) and reverse transcribed into cDNA using random hexamers (2.6.3).

2.10.3 Snapshot and Analysis

Snapshot protocol was performed with genomic DNA and cDNA samples in parallel as described in 2.8. Data was analysed using Gene Marker V2 2.0 demo software (Soft Genetics LLC). Ratios of the peak heights for two alleles were determined separately for genomic DNA and cDNA in test samples and controls. Allele ratio of cDNA was then normalized to allele ratio of genomic DNA.

3 ERYTHROID CULTURES

3.1 Introduction

3.1.1 Erythropoiesis

Erythropoiesis defines the process of differentiation and proliferation from hematopoietic stem cells (HSCs) to mature red blood cells (RBCs).

Erythrocytes (red blood cells) are derived from the myeloid progenitor cells. Committed erythroid progenitors referred to as burst forming unit-erythroid (BFU-e's) are the most immature haematopoietic cells that are already committed to the erythroid lineage. BFU-e develops into colony forming units-erythroid (CFU-e). CFU-e will differentiate into the first morphologically identifiable cell of the erythrocyte lineage, the pro-erythroblast, followed by successive differentiation into the basophilic erythroblast, the polychromatophilic erythroblast, and the acidophilic erythroblast, which is the last nucleated cell of the mammalian erythrocyte lineage (Figure 3.1).

During mammalian development, erythropoiesis occurs successively in the yolk sac, the foetal liver and the bone marrow. Erythropoiesis is composed of commitment and differentiation steps that restrict the differentiation potential and the proliferative capacity of the cells as they go through the erythroid-specific program of gene expression. Combinations of growth, microenvironmental and trans-acting factors that have crucial roles in promoting survival of erythroid progenitors regulate erythropoiesis. At the molecular level, trans-acting factors regulate transcription of genes involved in the regulation of the delicate balance between the opposing effects of proliferation promoting factors maintaining the renewal capacity of immature erythroid progenitors and differentiation-inducing factors required for successful terminal maturation of erythroid progenitors into red blood cells (Figure 3.1) (Bauer et al 1999).

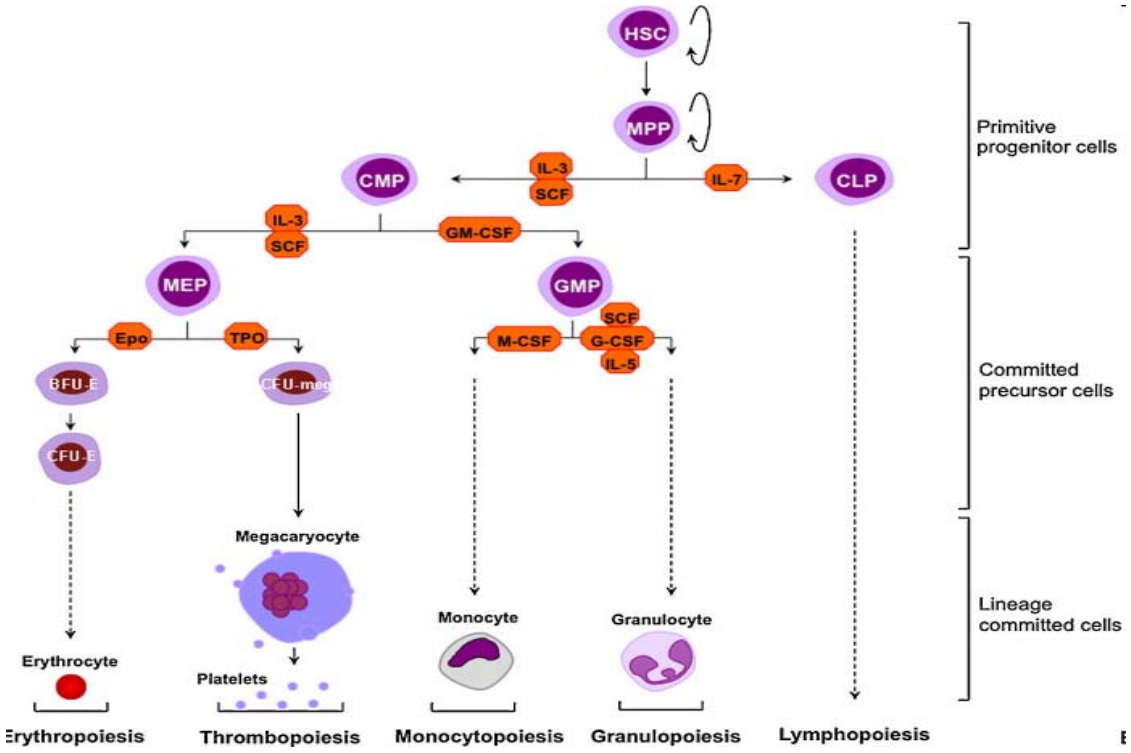


Figure3.1: Cytokines are involved in lineage commitment during haematopoiesis. Figure adopted from (Bauer et al).

3.1.2 Erythroid Cell Cultures

The study of human erythropoiesis *in vitro* requires a robust culture system that consistently and reliably generates large numbers of immature erythroblasts that can be induced to differentiate synchronously. The conditions of culture systems are developed to mimic the natural environment of erythropoiesis. The cells yielded using these culture systems are very similar to their *in vivo* counterparts. This means morphological changes and gene expression changes accompanying erythroid differentiation can be studied in these cells.*In vitro* models of human erythropoiesis are useful tools in studying normal and pathologic erythroid maturation, in addressing the pathophysiological mechanisms underlying RBC diseases and in testing the effects of various pharmacological agents.

They are also useful at providing us with cells at different stages of maturation for analysis of the different transacting factors that are active at different stages of erythropoiesis.

Erythroid progenitor cells (CD34+) can be isolated from various sources including bone marrow aspirates, fetal livers, cord blood and peripheral blood buffy coats. CD34 is a cell surface glycoprotein and functions as a cell-cell adhesion factor which is expressed in erythroid progenitor cells. There are also commercially available CD34+ cells that are obtained by cell sorting.

In our erythroid cell cultures, we use peripheral blood as the source. Peripheral blood buffy coat displays two main advantages:

- Easy accessibility
- homogenous population of BFUe meaning a uniform starting point for culturesLeberbauer et alalso suggested that cell differentiation is more synchronised with samples from peripheral blood compared to cells sorted for CD34+ (Fibach et al 1989 and Leberbauer et al 2005).

The isolated cells can be grown in different types of assay media. Semi-solid medium colony forming assays were commonly used but they have limiting characteristics such as immobility of the cells which limits cell interactions and continuity of the culture which limits addition of different factors at different stages. A two-phase liquid culture was applied by Fibach and colleagues to overcome these limitations. The group proposed a two-phase liquid culture system that supports the growth and maturation of human erythroid progenitor cells. The original Fibach method was modified by Jackie Sloan-Stanley (WIMM, Oxford) and we applied her method with her generous help. Three different methods of liquid culture were performed during the project for isolation of erythroid progenitor cells. Two methods (two modified Fibach methods) consists of two phases while the third method has three phases in which Phase II of the first two methods is separated as phase II and phase III. Results presented in this chapter used erythroid progenitor cells derived from the modified three-phased culture method; but other cultures method will be discussed for comparison. We needed cells at early stages i.e. (proerythroblasts, basophilic normablasts) so the cells were not transferred

to phase III as differentiation was not required for our study. Table 3.1 compares key steps of three different methods.

Key Points	2-Phase Fibach Culture	2-Phase Stem-span Culture	3-Phase Culture
Plastic	Cells are incubated in flasks.	Cells are incubated in flasks.	Cells are incubated in petri dishes or 6 well plates for better aeration.
Base Medium	Biological Foetal Calf Serum(FCS) is used as base medium. Batch variable	Semi-synthetic Stemspan is used as base medium. Batch variable	Stemspan is used as base medium. Batch variable
Cytokines	Conditioned medium includes key cytokines needed for proliferation of early progenitor cells.	IL3 replaces conditioned medium for proliferation of early progenitor cells.	Additional cytokines (lipids, IGF1) are included in phase I and phase II to induce proliferation
Media volume	10-25 ml depending on cell counts	10-25 ml depending on cell counts	2-10 ml depending on cell counts
Concentration of cells	Cell concentration is 1million cells /ml during phase I	Cell concentration is 1million cells /ml during phase I	Cell concentration is 5-10million cells /ml during Phase I leading to induced cell interactions
T-lymphocyte depletion	Cyclosporin is used to kill T-lymphocytes	Cyclosporin is used to kill T-lymphocytes	Cyclosporin is not used to kill T-lymphocytes. Histopaque layering can be performed to deplete T-lymphocytes
Macrophage depletion	Macrophages are depleted during phase 1 and 2 by changing the flasks	Macrophages are depleted during phase 1 and 2 by changing the flasks	Macrophages are depleted during phase 1 and 2 by changing the petri dishes

Table 3-1: Comparison of three erythroid culture methods

3.1.2.1 Two-Phase Fibach Culture

50 mls of peripheral blood in EDTA was diluted 1:1 with PBS. This was layered onto a density gradient centrifugation liquid- Histopaque-Ficoll (Sigma-Aldrich, UK) and centrifuged to separate the blood components. The buffy coat containing lymphocytes and progenitor cells was collected and washed with PBS. The washes should be performed till the supernatant is transparent but there is fine balance between washing the cells and losing them so in the case of a small cell pellet fewer cell washes should be performed.

The cells were counted and cultured in flasks at 1-2 million/ml in Phase I medium [DMEM-Alpha modified medium (Sigma, Poole, UK), 5 U/ml penicillin, 5 µg/ml streptomycin, 1µg/ml cyclosporin A, 2mM L-Glutamine, 10% foetal calf serum (Sigma, Poole, UK) and 10% conditioned medium from cultures of 5637 bladder carcinoma cell line (ATCC)] at 37° in 5% CO_2 for 6 days.

Because of the biological background of Foetal calf serum each batch should be tested since some of the batches can inhibit cell growth. The concentration of cells is crucial for a successful cell culture since diluted cultures limits cell interactions and crowded cultures lead to apoptosis.

The first phase isolates erythroid progenitor cells by depleting all other cell types while expanding growth of the progenitor cells. Cyclosporin A was used in depleting lymphocytes. 24 hours later the entire culture suspension was moved to a fresh flask, this step removes adherent monocytes in the first flask. Monocytes are included in the buffy coat with the CD34+ cells during isolation. The default pathway of the early committed Myeloid/Erythroid pathway is towards monocytes. Excessive numbers of monocytes disrupt the homogeneity of the culture and alter proliferation/differentiation

patterns of the CD34+cells. On the other hand, complete depletion of monocytes in the culture is not advisable since monocytes/macrophages are essential sources of iron for the differentiating erythroblasts so they should be diluted by changing the flasks instead of being depleted by chemicals. After 6 days in phase I culture, the cells are centrifuged and resuspended in Phase II medium [DMEM alpha modified Medium, 5 U/ml penicillin, 5 µg/ml streptomycin, 1 U/ml human recombinant erythropoietin [EPO; Eprex UK] and 10 ng/ml SCF, 2mM L-Glutamine 30% Foetal calf serum , 10^{-5}M β-Mercaptoethanol, 10^{-6} mol/l dexamethasone ,0.3mg/ml Holotransferrin(ICN) , 1% Bovine serum albumin (Sigma, Poole, UK) and the cells are maintained at a concentration of 1 million/ml for up to 15 days. The second phase promotes the proliferation and maturation of the erythroid progenitor cells. This selection is achieved by administration of key growth factors including erythropoietin (EPO) and stem cell factor (SCF).

3.1.2.1 Two-Phase erythroid culture using Stem Span

This method is an adaption of the original Fibach method where biological Foetal calf serum is substituted with semi-synthetic Stem Span and cytokines such as IL3 and stem cell factor replaces the 5637 cells conditioned medium. Even though Stem span is semi-synthetic, it still contains biological reagents with batch variation so each batch should be tested before proceeding.

50 mls of peripheral blood in EDTA is centrifuged to separate the blood components. The buffy coat which contains the lymphocytes and progenitor cells at the yellow plasma interface is collected and layered onto a density gradient centrifugation liquid: Histopaque-Ficoll (Sigma-Aldrich, UK) and centrifuged to separate red and white blood cells. The lymphocytes and other mononuclear cells remain at the plasma-histopaque interface from where they are collected and washed in PBS. Cells are counted and

cultured in flasks at 1-2 million/ml in Phase I medium (Stem Span Medium [Stem Cell Technologies, Vancouver, Canada], 10 U/ml penicillin, 10 µg/ml streptomycin, 1µg/ml cyclosporin A, 25 ng/ml interleukin 3 [IL-3; Sigma, Poole, UK] and 50 ng/ml human stem cell factor [SCF; Sigma, Poole, UK]) at 37° in 5% CO_2.24 hours later the entire culture suspension is transferred to a fresh flask, this step removes adherent monocytes in the first flask.

After 7 days in phase I culture, cells are centrifuged and resuspended in Phase II medium (Stem Span Medium, 10 U/ml penicillin, 10 µg/ml streptomycin, 2 U/ml human recombinant erythropoietin [EPO; Sigma, Poole UK] and 50 ng/ml SCF) and maintained at a concentration of 1 million/ml for up to 20 days. The critical steps of the original two phase Fibach Culture are also valid in this method. The concentration of the cells and reduction of excess monocytes are crucial steps. Cyclosporin A was used to deplete lymphocytes as in the first method.

2.1.3 Three-Phase erythroid culture

This method was described by Van Den Akker et al (Haemotologica 2010) and kindly provided to our group. I have followed the protocol with few adaptations. The main advantage of the method is that more cells can be harvested at the end of the protocol allowing the use of smaller volume of blood.25 mls of peripheral blood in EDTA is diluted 1:1 with PBS and layered onto a liquid density gradient (Histopaque 1.077) (Sigma-Aldrich, UK) and centrifuged to separate the blood components. The buffy coat containing lymphocytes and progenitor cells is collected and washed with PBS.
Cells are counted and cultured in petri dishes at 5-10 million/ml in Phase I StemSpan medium containing: IL-3 – 1ng/ml, human recombinant erythropoietin [EPO; Eprex UK] – 2U/ml, Stem cell factor – 100ng/ml, Dexamethasone - 1µM, Insulin growth factor

1(IGF1) – 40ng/ml, Cholesterol rich mix 50μg/ml, 5 U/ml penicillin, 5 μg/ml streptomycin for 5 days.

After 6 days the cells are centrifuged and resuspended in Phase II StemSpan medium containing: human recombinant erythropoietin [EPO; Eprex UK] – 2U/ml, Stem cell factor – 40ng/ml, Dexamethasone - 1μM, Insulin growth factor 1(IGF1) – 40ng/ml, Cholesterol rich mix 50μg/ml, 5 U/ml penicillin, 5 μg/ml streptomycin for 7-25 days. Cells are cultured in petri dishes at 0.5-1.0 million/ml. Differentiation can be induced after a minimum of 7 days in phase II which involved collecting the cells and culturing them in phase III StemSpan medium containing:
Human recombinant erythropoietin [EPO; Eprex UK] – 10U/ml, Insulin growth factor 1(IGF1) –, 5 U/ml penicillin, 5 μg/ml streptomycin, Holotransferrin 1mg/ml,Insulin 10ng/ml,thyroid hormone 1μM , 3 %human serum (male) for 2-7 days.This method differs from the 2-phase methods described and has different critical steps (see Table 3.1 for summary of differences between the 3 methods). Briefly, in the 3-phase method the cells are cultured at very high concentrations in phase I in petri dishes or plates for a better aeration of the culture. If an excessive population of lymphocytes is observed, a histopaque layering is performed to remove the lymphocytes avoiding the use of cyclosporine- A in the medium. One critical point is to maintain the monocytes at a minimal level. The strategy to maintain low monocyte levels rely on the adherence property of the monocytes. The adherence of the activated monocytes to the surface of petri dishes is enabled by a specific coating (CELLSTAR surface treatment) of the plastic dish (Figure 3.2).

The media include cytokines such as insulin growth factor like 1 (IGF1) and cholesterol rich mix that were not present in pervious culture systems. Table 3.2 summarizes functions of cytokines used in phase I and phase II.In Phase III, the high dose of

erythropoietin in the medium increase the differentiation rate of erythroid cells which make it difficult to harvest them at specific stages.

For each ChIP experiment 10^6 cells were required which can easily be rich with this method. We assessed cell development based on the morphology analysis of cell cytospins. The purity and maturation of cells was further analysed by flow cytometry. The full methods are outlined in chapter 2.

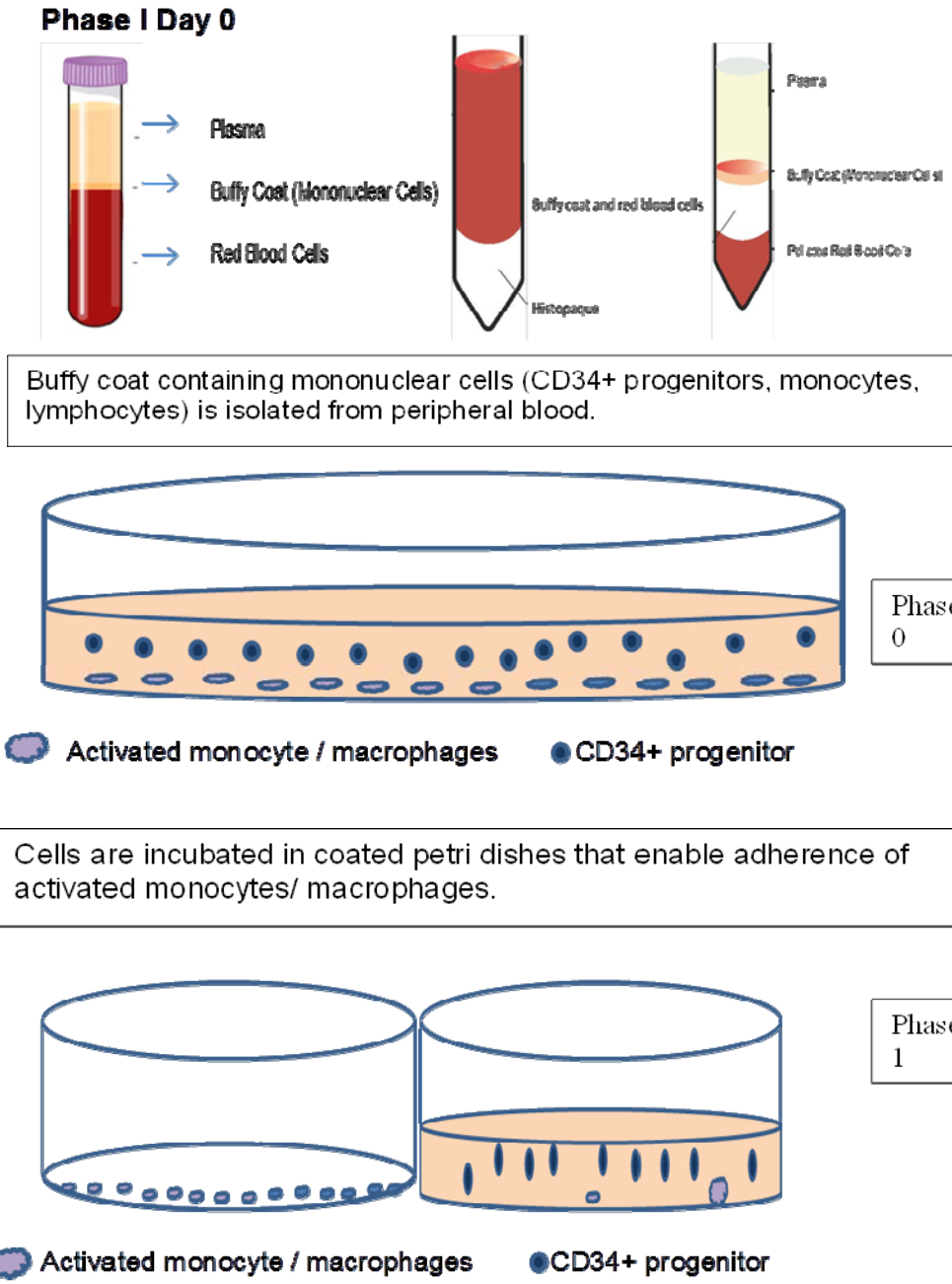


Figure 3.2: Monocyte / Macrophage depletion by petri dish changes according to method adopted from Van Den Akker et al (Haemotologica 2010).

Reagents	Final Conc.	Phase I	Phase II	Function
Stemspan	-	100ml	100ml	It is a serum free expansion medium specifically for culture of haemapoietic cells.
Erythropoietin (EPO)	2U/ml in phase 1/2 10U/ml in phase 3	200μl	200μl	EPO drives the proliferation and expansion of the developing erythroid progenitors from BFU-e stage through to terminal differentiation.
Interleukin 3 (IL3)	1ng/ml	100μl	-	IL-3 is a cytokine that promotes cell survival and proliferation of HSCs and early progenitor cells.
Stem Cell Factor (SCF)	100ng/ml	400μl	400μl	Induces cell survival and proliferation of early erythroid progenitors and prevent apoptosis at later stages.
Dexamethasone (Dex)	1μM	100μl	100μl	It activates proliferation of immature erythroid progenitor cells and delay differentiation.
Insulin growth factor like 1 (IGF1)	40ng/ml	100μl	100μl	IGF1 increase proliferation of early erythroid progenitors and prevent apoptosis.
Penicillin &Streptomycin	1:100	1ml	1ml	Prevent bacterial contamination
Cholesterol rich mix	50μg/ml	500μl	500μl	Induces cell survival and proliferation of early erythroid progenitors.
L-Glutamine	1:100	1ml	1ml	Glutamine supports the growth of cells that have high energy demands

Table3-2: Summary of cytokine function used in erythroid culture

3.2 Results

Maturity and differentiation stage of the erythroblasts were monitored by morphology in cytospins and expression of cell surface antigens by FACS.

3.2.1 Analysis of cell morphology

During Phase II, the cells are maintained at a concentration $1*10^6$/ml to allow proliferation of erythroid progenitors. Cell differentiation was monitored daily by cell cytospins (see chapter 2 for full methods).

Small round erythroid progenitor cells were visible in the culture from day 0 of Phase II with the presence of some prepro-erythroblasts which increased in number by day 2. The cell cytospin taken on day 4 showed the majority of the cells in the culture were now pro-erythroblasts. As the cells developed into polychromatic erythroblasts (day 7) they reduce in size and the nuclei become more condensed. By day 9 the cells were smaller yet with a pale cytoplasm as the dark nuclei move to one side of the cell known as orthochromatic erythroblasts. Soon the nuclei were expelled from the cells as seen at day 13 (Figure 3.3).

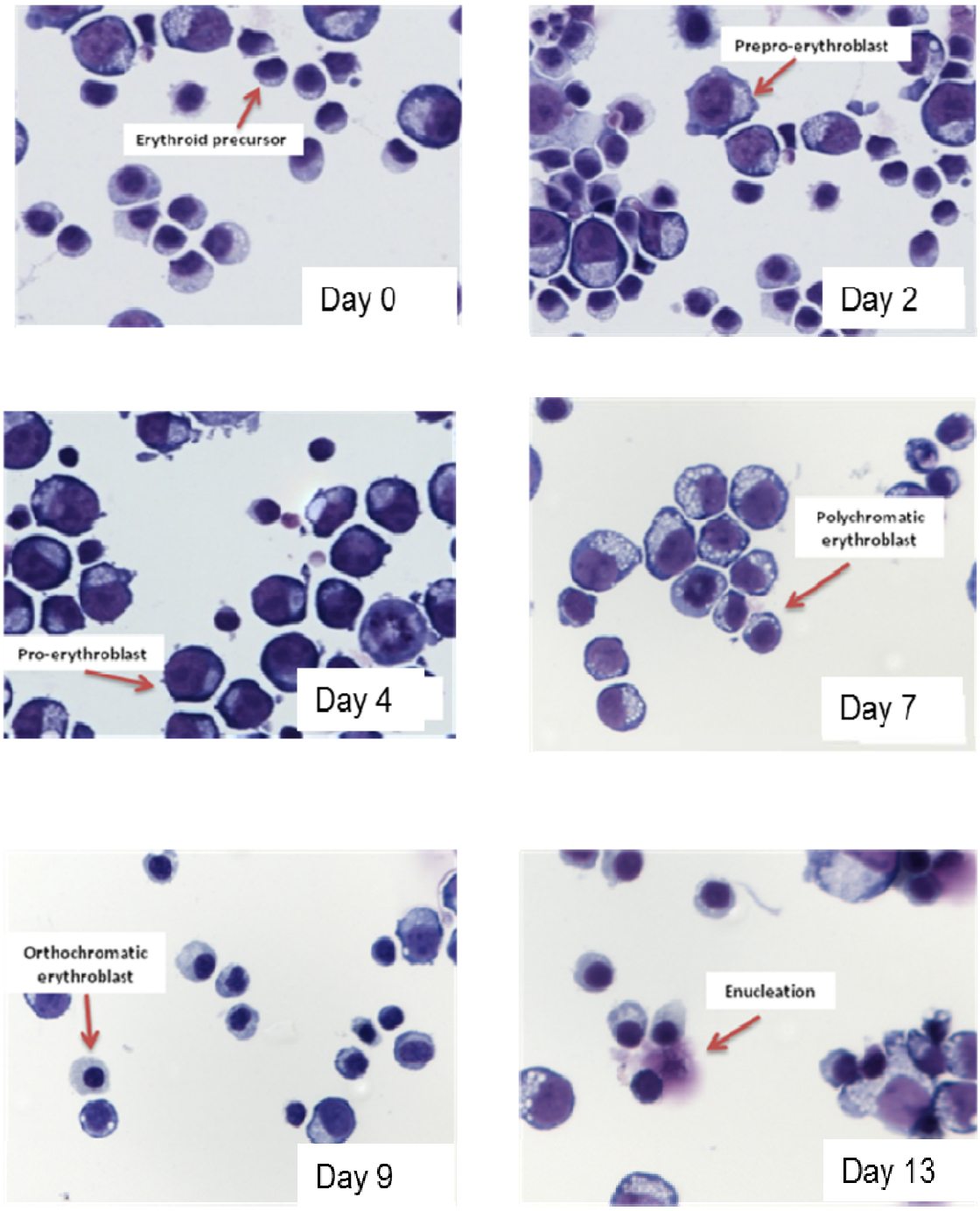


Figure3.3: Differentiation of primary human erythroid progenitor cells.
Cytospins of cells taken in Phase II day 0-13 of the culture system adapted from Van Den Akker et al (Haemotologica 2010). Morphological stages of cells are marked out on each day.

3.2.2 Expansion of erythroid progenitor cells in Phase I of culture

We isolate erythroid progenitors (CD34+) from peripheral blood buffy coats containing mononuclear cells such as lymphocytes, monocytes and progenitor cells. The first phase of culture isolates erythroid progenitor cells and allows their growth by depleting all other cell types.

Expansion of erythroid progenitor cells was monitored by fluorescence activated cell sorting (FACS). On the day of isolation (day0) the percentage of erythroid progenitors (CD34+) was less than 10 %. After two days in phase I majority of adherent cells (monocytes) were depleted by frequent petri dish changes leading to an increase in the percentage of CD34+ cell: 18 % at day 2 and 32 % CD34+ cells at day 5 (Figure 3.4).

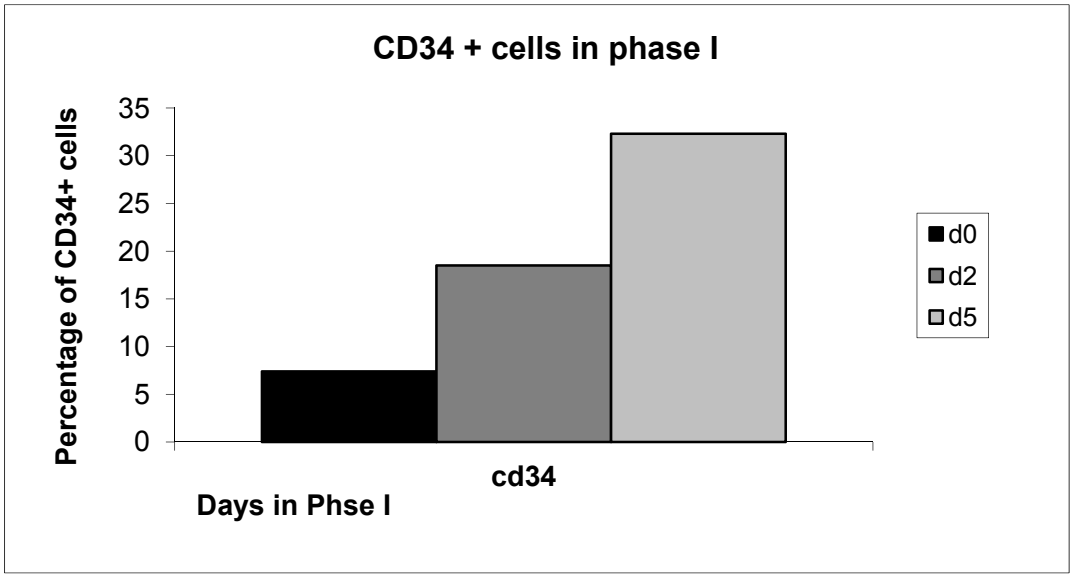


Figure 3.4: Percentage of CD34+ progenitor cells during primary human erythroid progenitor cell cultures by FACS analysis.
The percentage of CD34+ cells in Phase I day 0, day 2, and day 5 of the modified three-phase culture are obtained using anti-CD34+ antibody.

3.2.3 Purity and differentiation of erythroid cells in Phase II of culture

We used two human erythroid cell surface proteins to monitor cell development and purity. Glycophorin A (GPA) is an erythroid specific protein and is up-regulated as the cells develop into erythrocytes. Transferrin receptor protein 1 (CD71) (Furthmayr et al 1975) is highly expressed on the surface of all proliferating haematopoietic cells including erythroid progenitors during the early stages of development.

At early stages of culture (day 4) cells start to proliferate, 82 % of the cells express CD71 but only very few express GPA (Figure 3.5).

At phase II day 9 of the culture, CD71 expression on cells is high (82% on day 4 to 95% on day 9), the maximum that it would reach.
GPA is expressed later as the cells differentiate down the erythropoietic pathway. We see a rapid increase in GPA expression from day 4, when 14% of cells analysed are positive to 73% on day 11 when the highest level of GPA positive cells is observed. GPA positive, CD71 negative cells would appear if further differentiation of the cells is induced by transferring the cells to phase three (Figure 3.5).

3.2.4 Gene expression profiles

For ChIP experiments, we need to harvest cells at the peak expression stage for the transcription factor of interest. We constructed expression profiles of the different genes by RT-PCR of RNA extracted from erythroid cells at sequential days of Phase II. Ct values were normalised to the house-keeping gene HPRT. Figure 3.6 presents expression *MYB, KLF1* and *GATA-1* in correlation with FACS plots and cytospins.

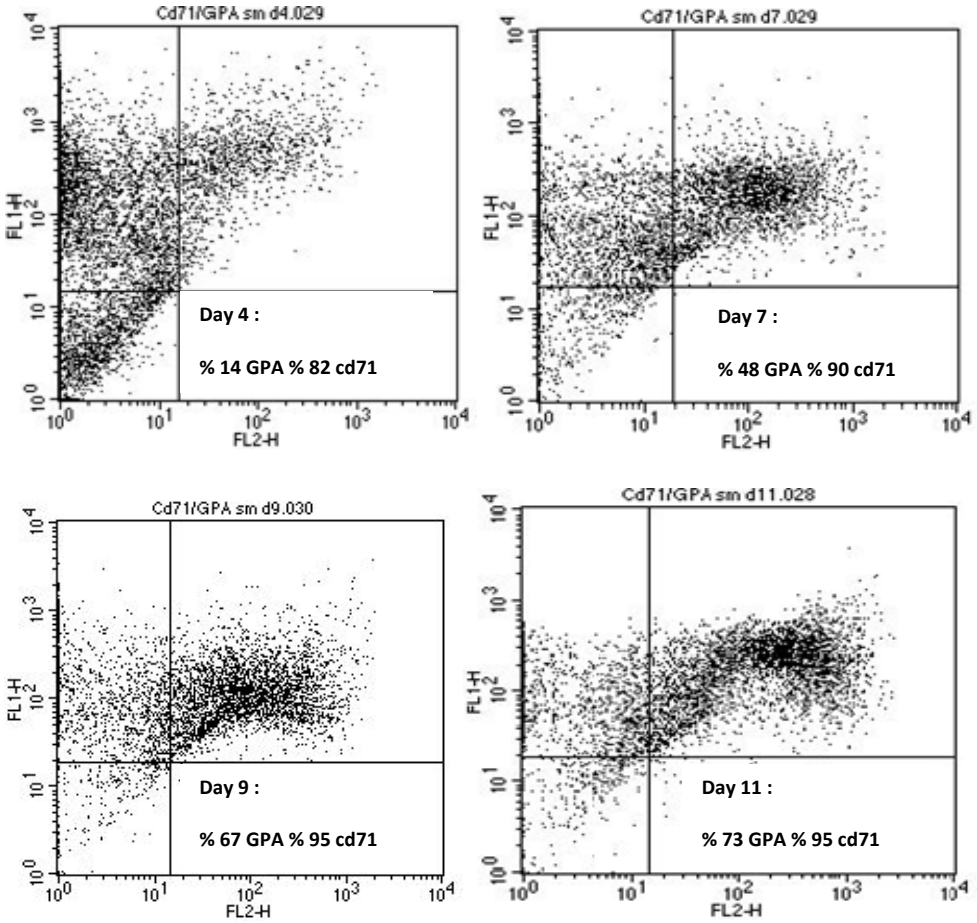
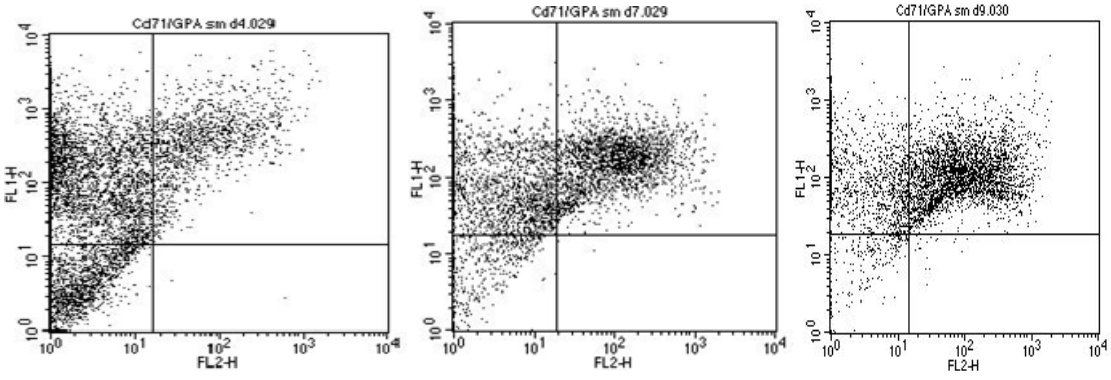


Figure 3.5: Differentiation and purity of primary human erythroid progenitor cells by FACS analysis.
Cells in Phase II day 4-11 of modified three-phase culture are double stained with antibodies against CD71 and GPA. The y-axis shows CD71 expression and the x-axis shows GPA expression. Lower left box: double negative cells. Upper left box: CD71 positive, GPA negative cells. Upper right box: GPA and CD71 positive cells. Lower right box: GPA positive, CD71 negative cells. Markers (black lines) separating each field were set according to single stained and unstained control samples from the same culture at each day.



Day 4:

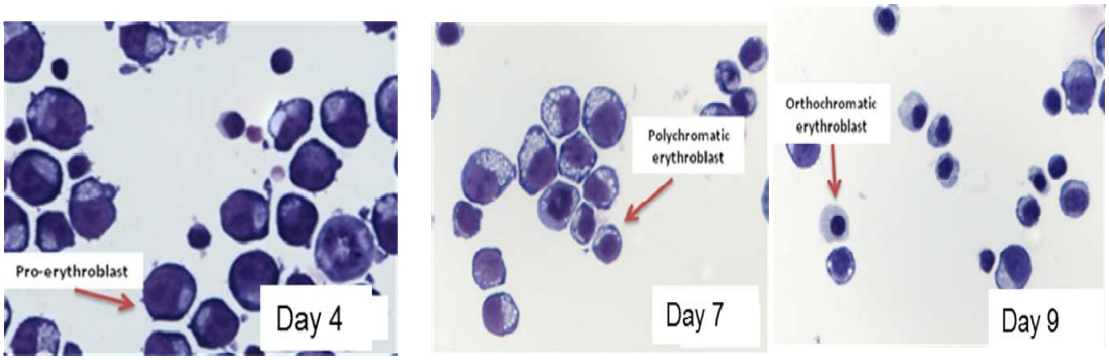
% 14 GPA % 82 cd71

Day 7:

% 48 GPA % 90 cd71

Day 9:

% 67 GPA % 95 cd71



Expression levels

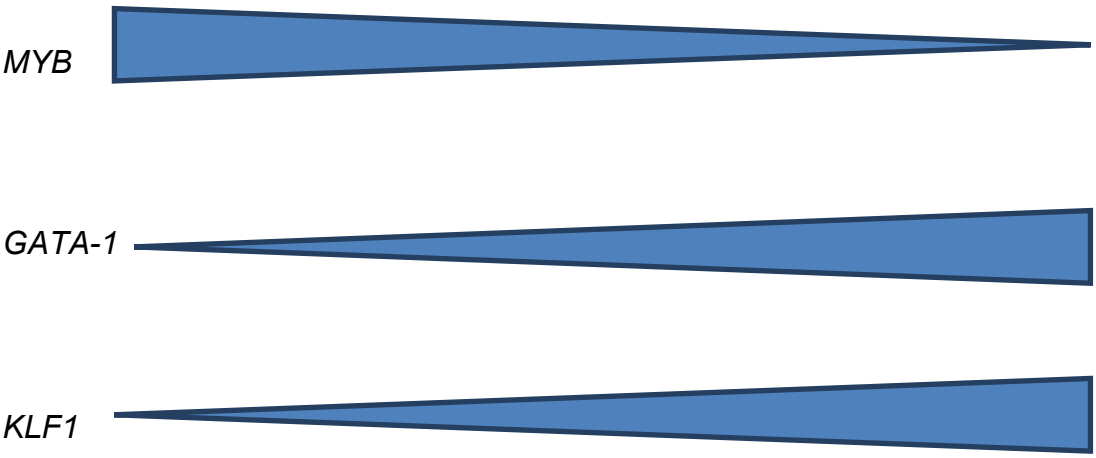


Figure 3.6: Expression profiles of key transcription factors during erythroid differentiation.

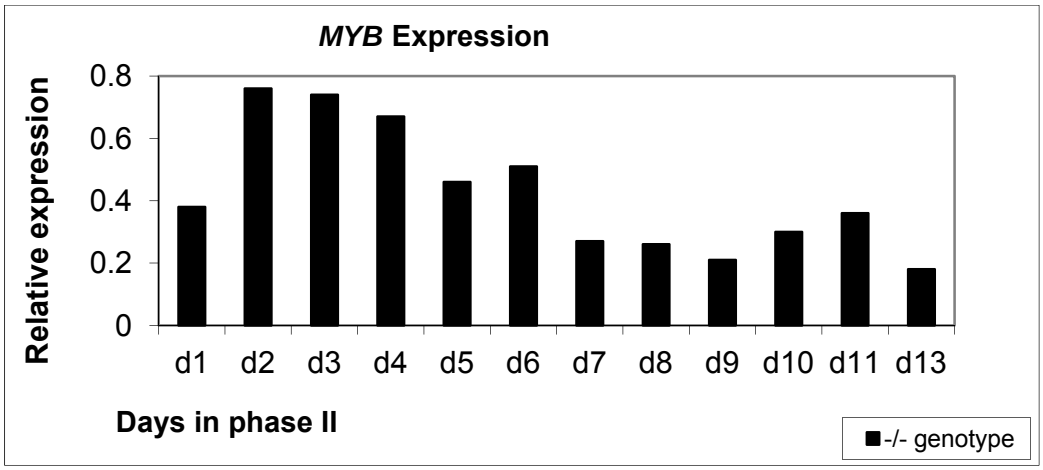
MYB was highly expressed at early stages of erythroid culture (day2-day4) when majority of the cells are pro-erythroblasts. As cells differentiate becoming basophilic or polychromatic erythroblasts *MYB* expression is down regulated remaining at a low level after day 7 (Figure 3.7).

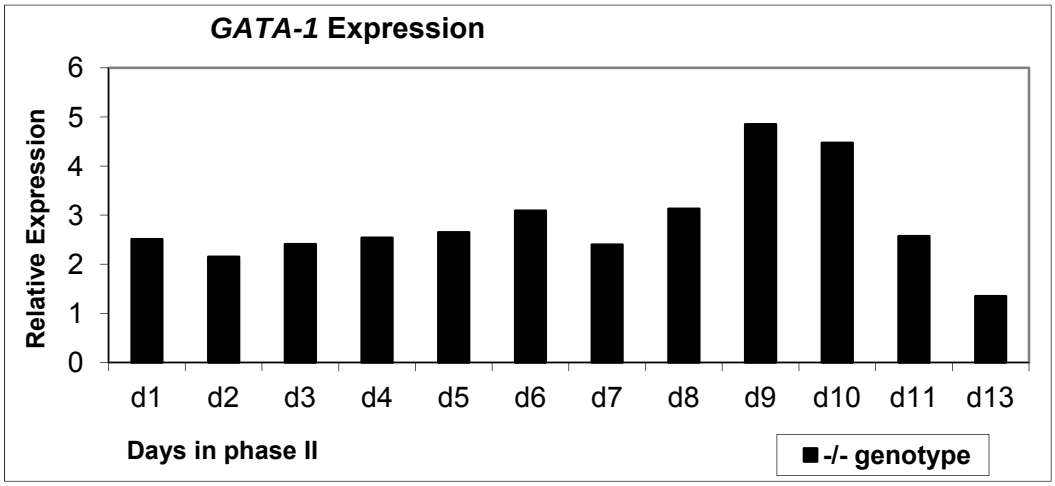
GATA-1 was up-regulated as cells developed, reaching the peak expression on day 9. This coincides with the stage where all cells are erythroblasts and GPA is highly expressed. As the cells further differentiate into orthochromatic erythroblasts and start enucleating GATA-1 expression is down-regulated (Figure 3.7).

KLF1 expression follows a similar pattern to that of *GATA-1* expression, reaching a peak on day10. The main difference from*GATA-1*expression is that *KLF1* is not down-regulated as much as *GATA-1* around day 13 (Figure 3.7).

This data shows that two genes (*KLF1 andGATA1)* of interest are at peak expression around day 9 and 10 of culture in phase II. This coincides with the high levels of GPA expressing cells (~70%, Figure 3.5) at the polychromatic erythroblast stage (Figure 3.6).

On the other hand cells should be harvested at early stages (day 4) for *MYB* related experiments when majority of the cells are highly proliferating thus expressing high levels of *CD71.*





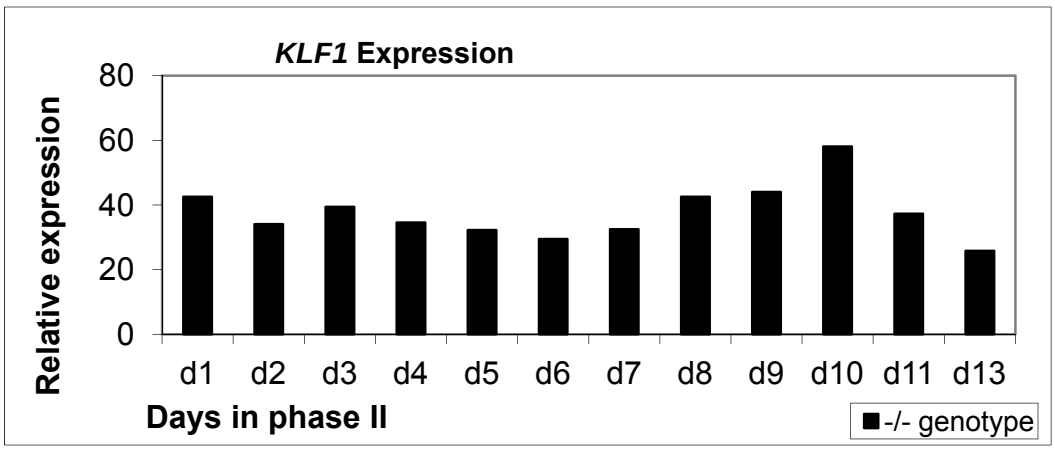


Figure 3.7: Relative expression profile of *MYB, GATA-1* and *KLF1* in erythroid progenitors from phase II day1-13.
Results are analyzed by RT-PCR in duplicate using gene specific primers on erythroid RNA from a single erythroid culture. Individual designated as -/- is homozygous for the absence of HMIP block 2 SNPs associated with elevated HbF. Data were normalized to housekeeping gene HPRT.These results are from erythroid culture material of one individual.

3.2.5 Erythroid Culture Kinetics of individuals with different genotypes

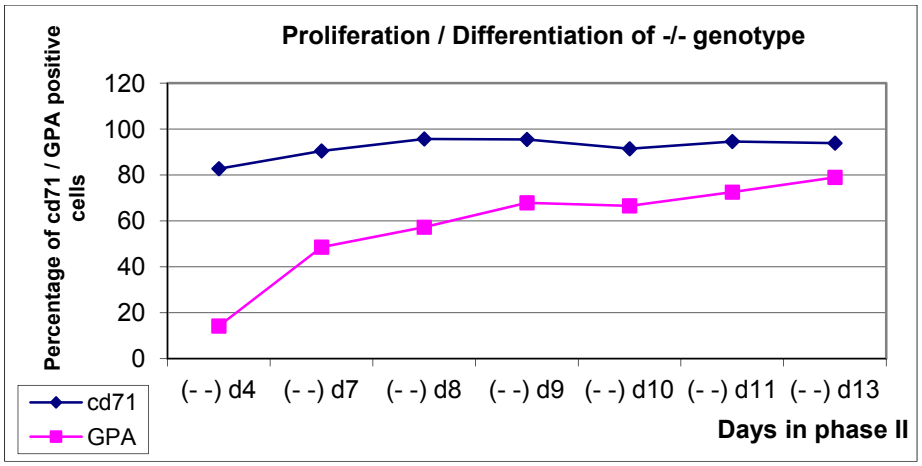
For the next objective, we wanted to obtain erythroid progenitor cells from individuals that are either homozygous (+/+) for the haplotype associated with high HbF across the 6q QTL (blocks 1, 2 and 3) or homozygous for the absence of the SNPs associated with increased HbF levels (-/-). Our initial objective was to investigate differences in the erythroid culture kinetics between these individuals.

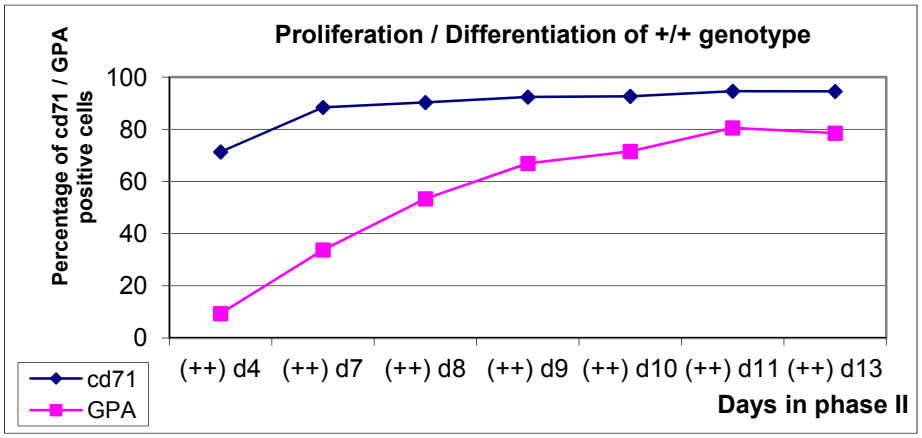
We used two human erythroid cell surface proteins: Glycophorin A (GPA) and transferrin receptor protein 1 (CD71) to monitor proliferation and differentiation of erythroid progenitors.

The results suggested that differentiation of erythroblasts was accelerated in +/+ individuals. In a +/+ individual percentage of GPA positive cells was 9% at day 4 and it reaches 80 % in 7 days at day 11 compared to that for a -/- individual where percentage of GPA positive cells reaches 72 % at day 11 from 14 % at day 4 (Figure 3.8 A). The difference became striking when results from each individual were corrected for the percentages in CD71 positive cells and then normalised to day 4 of each individual. Figure 3.8 B demonstrates relative increase of GPA positive cells intraindividually confirming the accelerated increase of GPA positive cells in the +/+ individual.

The differential increase in monocytes between the two genotypically different individuals (-/- and +/+) was also monitored using cell surface antigen CD14, a specific marker for monocytes. The results suggested an increased number of CD14 positive cells in +/+ individuals at two stages of phase II (day4 and day11). The increase in monopoiesis can be the result of inefficient commitment of early progenitors to erythroid pathway since cells that are not committed to erythroid lineage commits to monocytic lineage by default (Figure 3.9).

A





B

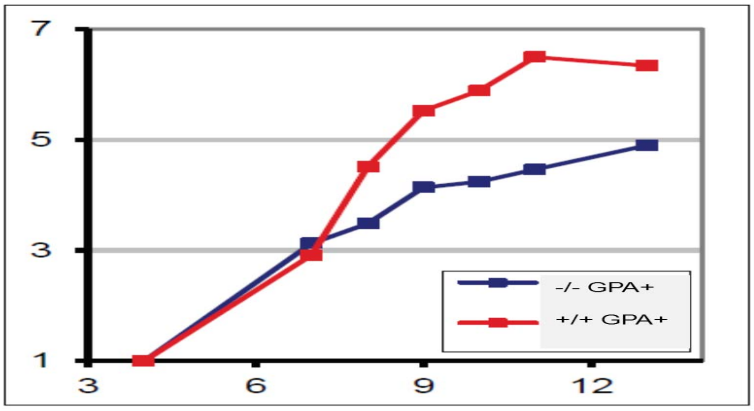


Figure 3.8: Differentiation and proliferation curves from +/+ and -/- individuals.
A: Differentiation and proliferation kinetics of primary human erythroid progenitor cells was analysed by FACS analysis. Cells in Phase II day 4-13 of modified three-phased culture are double stained with antibodies against CD71 and GPA. B: Relative increase of GPA+ cells corrected any changes in CD71 % and then normalized to day 4 of each individual.These results are from erythroid culture material of two paired individuals.

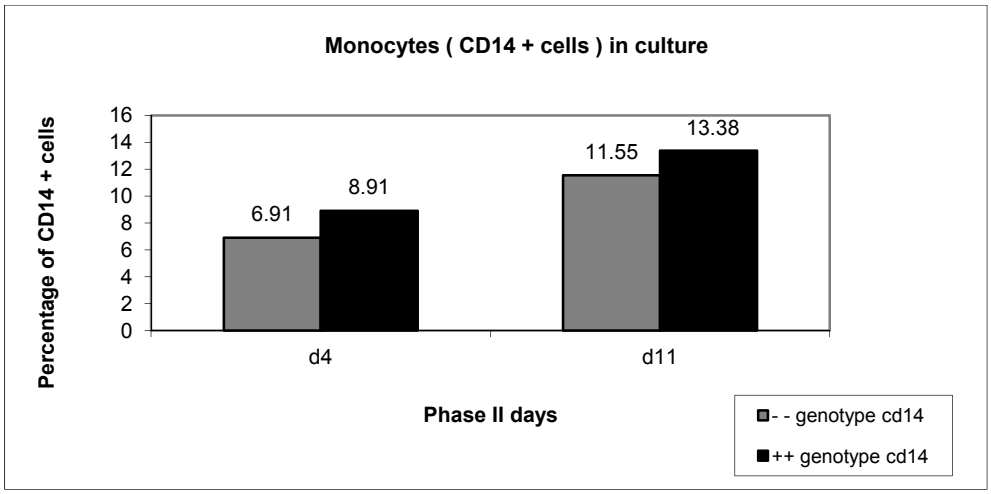
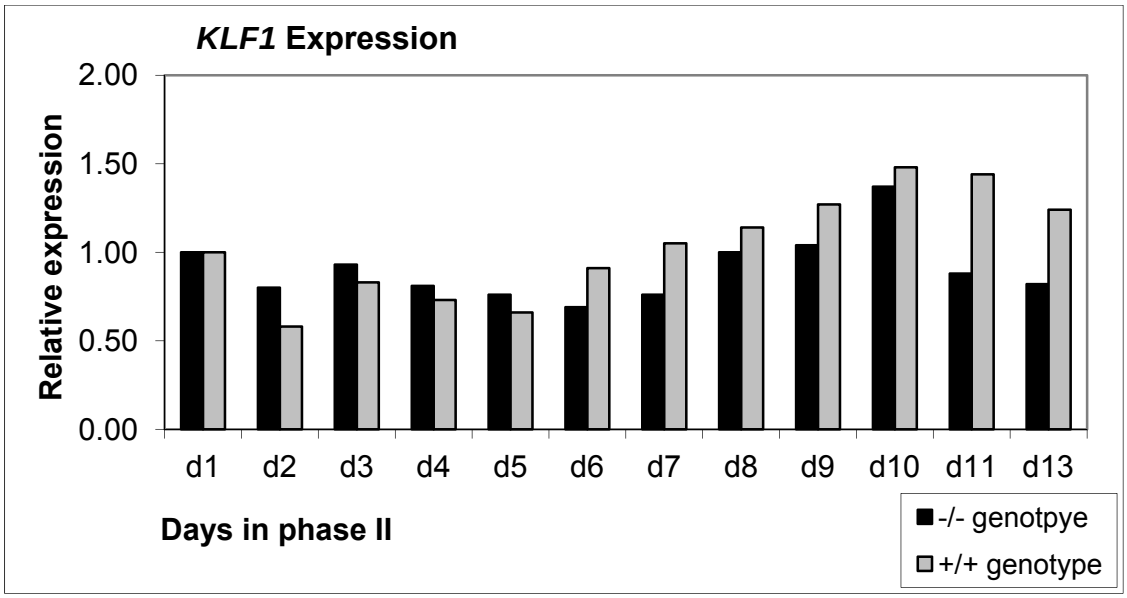
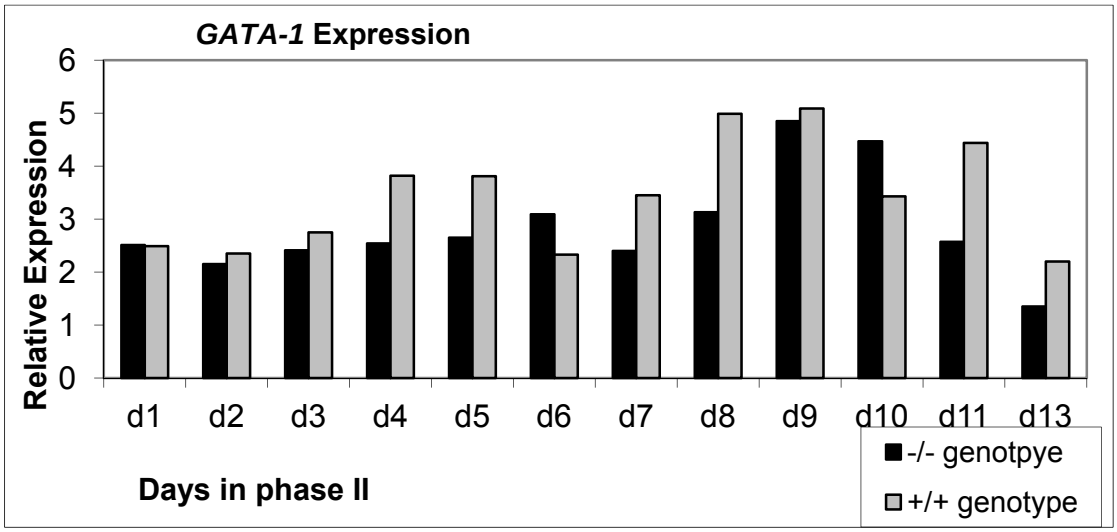


Figure3.9: Percentage of monocytes in primary human erythroid progenitor cell cultures is monitored by FACS analysis using an antibody against cell surface antigen CD14, which is a specific marker for monocytes and macrophages.

Our second objective was to harvest erythroid progenitors for ChIP experiments to investigate altered transcription factor binding. To ensure that we harvest the erythroid progenitor cells from the +/+ and -/- individuals at equivalent stages of gene expression we initially obtained a profile of gene expression in cells from individuals of both genotypes. Expression profiles suggested that day 9 or 10 would be optimal for *GATA-1* and *KLF1* when they are at their peak of expression. In the case of *MYB,* day 4 would be optimal as it is the peak of expression. Cells cultured from individuals of both genotypes shared a similar expression pattern in terms of peak of expression for *MYB* gene. *MYB* expression peaks at early stages of phase II than gradually decrease with both genotypes but +/+ individuals had a different pattern of *MYB* expression at later stages of erythroid culture (day 5–day 10) when compared to that of -/- individuals. For validation the cells were also closely monitored by FACS analysis as previously described. In performing ChIP on erythroid progenitor cells at similar stages of expression we ensure any difference observed in transcription factor binding is due to the alteration of binding by the SNPs and not because of differential expression of transcription factors (Figure 3.10).





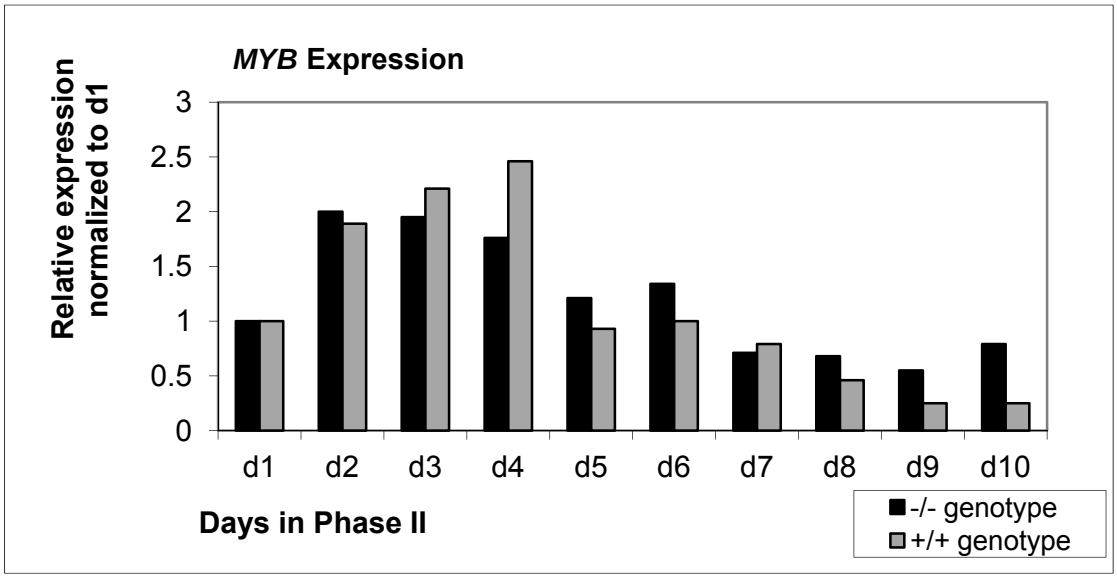


Figure 3.10: Relative *MYB, GATA-1* and *KLF1* expression profile from phase II.

3.3 Discussion

The functional experiments of the project depend on material from primary erythroid progenitors. Three different methods of liquid culture were performed during the project for isolation of erythroid progenitor cells in order to investigate the best culture method to obtain large number of homogenous and synchronized cells for expression and transcription factor binding studies.

Initial cell culture experiments were performed using Stem Span in the two-phase erythroid culture system, an adaption of the original Fibach method. Here the biological Foetal calf serum is substituted by semi-synthetic Stem Span and cytokines such as IL3 and stem cell factor replaces the 5637 cells conditioned medium.

Even though Stem Span is semi-synthetic, batch variation was observed and that led to a challenging time with the cultures i.e. some batches were inhibiting erythropoiesis and promoting monopoiesis some of our cultures, monocyte were as high as 50 % in the cultures. Despite the frequent flask changes it was difficult to deplete monocytes and macrophages during phase I as monocytes and macrophages that differentiate from progenitor cells are not activated and do not adhere to the flasks. Previous PhD student Karin Wahlberg tried depleting monocytes by use of a lysosomotropic agent, l-leucin methyl ester (LME), which causes selective death of monocytes in culture. LME treated cultures contained no macrophages on day 0 of phase II. A high amount of cell death was observed from day 6 of phase II with no viable cells visible on cytospins at day 10. In contrast, the control culture (no LME) had a small number of macrophages on day 0 of phase II and erythroid precursor cells had developed to a homogenous population of pro-erythroblasts at day 6.

This led us to conclude that limited numbers of macrophages are beneficial to erythroid cell development *in vivo* and *in vitro* preventing us from completely depleting them (Chasis & Mohandas 2008, Leimberg et al2005).

Another problem with cultures containing higher numbers of macrophages was accelerated differentiation. We have observed formation of erythroblast islands around macrophages that differentiate faster. It is known that macrophages are involved in erythroid development *in vivo* in the bone marrow by interacting with erythroid precursors in so called erythroblastic island consisting of a central macrophage surrounded by erythroblasts (Chasis and Mohandas, 2008). *In vitro* studies have demonstrated that macrophages promote proliferation and maturation of erythroid precursors in serum free cultures (Leimberg et al 2005) and macrophages have been shown to express erythroid potentiating cytokines (Bergamaschi et al 1993) including erythropoietin (Vogt et al 1989).

Because of the challenges of this culture method we decided to try the Fibach method that was modified by Jackie Sloan-Stanley (WIMM, Oxford). We applied her method with her generous help. Successful cultures were performed with this culture system but the major limitation was the limited number of cells that we could harvest. The lack of additional cytokines in the method limited proliferation of erythroblasts. For the purpose of ChIP experiments we needed a large number of homogenous populations of primary human erythroid cells at synchronised stages of expression.

Due to the limited number of cells we obtained using the Fibach method we had to increase the number of starting cells i.e. the amount of peripheral blood. Another challenge was the biological background of foetal calf serum. Each batch need to be batch tested since some of the batches could inhibit cell growth.

The third culture method that was applied in the project was described by Van Den Akker et al (Haemotologica 2010). I have followed the protocol with some small adaptations. This method is different from the traditional Fibach method as mentioned in the previous section. The main advantage is that more cells can be harvested at the end of the protocol allowing the use of a smaller volume of blood.

In the two-phase methods, Phase II includes reagents that favour both proliferation and differentiation of erythroid cells. In the three-phase culture, Phase II favours proliferation mainly where cells expand with limited differentiation due to the low concentration of erythropoietin while Phase III favours differentiation of erythroid cells due to the presence of tenfold more erythropoietin. Cell proliferation is limited at this stage due to lack of stem cell factor in the Phase III medium.

The three-phase culture system yields better cell numbers due to two main reasons. The method takes advantage of CD34- cells becoming CD34+ and commitment to the erythroid lineage during phase I. Apart for the small number of CD34+ progenitors, the mononuclear fraction contains mainly CD34- cells, monocytes and lymphocytes. Two main approaches have been used to enrich the fraction of progenitors over other cell types. In the 2-phase system (Fibach et al 1989;Pope et al 2000) mononuclear cells are grown in culture conditions that favour expansion of the progenitor cells followed by the depletion of contaminating monocytes and lymphocytes. The remaining progenitors are grown in a second phase in the presence of EPO resulting in erythroid proliferation and differentiation. The other approach of enriching progenitors selects the progenitors in the mononuclear cell fraction using antibodies for the surface marker CD34 (Keller et al 2006;Ronzoni et al 2008). Purified progenitors are cultured directly in the presence of Epo. The phase one of the three-phased method includes Epo and increased concentration of IL3 in order to direct CD34- cell to become CD34+. The cell interactions are crucial for this stage so cells are kept at high concentrations.

We observed significant increase in the proportion of CD34+ cells at the end of phase one (Figure 3.4). The second reason is the presence of the additional cytokines that induce proliferation (Table 3.2).

Jiang et al suggested that individuals with elevated HbF levels had accelerated differentiation and relatively lower levels of *MYB* expression compared to individuals with normal or low HbF. We wanted to substantiate these findings with our culture

system with primary erythroid cultures from individuals that are either homozygous (+/+) for the haplotype associated with high HbF across the 6q QTL (blocks 1, 2 and 3) or homozygous for the absence of the SNPs associated with increased HbF levels (-/-).

Our results confirm Jiang's initial observations of accelerated differentiation and relatively increased monopoiesis in +/+ individuals that have higher HbF levels (Figure 3.8). One possible explanation for the accelerated differentiation is the difference in*MYB* expression profiles between these individuals. The commitment of haematopoietic stem cells to the erythroid lineage is dependent on distinct threshold levels of MYB. Reduced MYB levels favours commitment of haematopoietic progenitors to the megakaryocytic (platelet precursor) and monocytic lineages over the erythroid lineage (Emambokus et al 2003;Lu et al 2008). Consequently, accelerated erythropoiesis as a response to relatively low MYB levels generates prematurely terminated erythroid cells synthesising more HbF (Jiang et al 2006). Thus, modulations in MYB levels could well underlie the variation in erythroid, platelet and monocyte counts as well as HbF levels associated with *HMIP*.

Increased number of monocytes present with +/+ individuals can also promote accelerated maturation of erythroid precursors. It is known that macrophages are involved in erythroid development *in vivo* in the bone marrow by interacting with erythroid precursors in so called erythroblastic island consisting of a central macrophage surrounded by erythroblasts (Chasis and Mohandas, 2008).

In vitro studies have demonstrated that macrophages promote proliferation and maturation of erythroid precursors in serum free cultures (Leimberg et al 2005) and macrophages have been shown to express erythroid potentiating cytokines (Bergamaschi et al 1993) including erythropoietin (Vogt et al 1989).

Our second objective was to harvest erythroid material for transcription factor binding experiments to investigate altered transcription factor binding. To ensure that we harvest the erythroid progenitor cells from the +/+ and -/- individuals at equivalent

stages of gene expression we initially obtained a profile of gene expression in cells from individuals of both genotypes. Expression profiles suggested that day 9 or 10 would be optimal since *GATA-1* and *KLF1* are highly expressed at these days. Also there is no significant difference between individuals of different genotypes in terms of expression. As validation the cells were also closely monitored by FACS analysis as previously described. In performing ChIP on erythroid progenitor cells at similar stages of expression we ensure any difference observed in transcription factor binding is due to the alteration of binding by the SNPs and not because of differential expression of transcription factors (Figure 3.9).

Performing primary erythroid culture experiments proved to be challenging. I have noted some of the crucial steps for fellow colleagues during optimizations stages:

- Reagents with biological backgrounds should be batch tested.
- Most reagents are fragile and should be kept in frigde once thawed; yet the reagents should not be kept in the fridge for more than a month.
- Media volumes are also critical and evaporation of medium should not be ignored when media volumes are decided.
- Cell concentrations vary accorrding to the phase of culture. Cells are incubated at high cell concentrations at phase I to trigger cell interactions while cell concentration should be lower in phase II to enable efficient prolifereation.

I have decided to use the three phase culture system after trying all three culture systems because of the following advantages:

- This system yields better cell numbers allowing the use of a smaller volume of blood and the harvest of the cells at different stages.
- Introduction of EPO at phase I induces proliferation leading to a shorter time period of culturing since cells reach the appropriate stage faster.

4 BIOINFORMATICS

4.1 Introduction

The chromosome 6q locus consists of a large number of single nucleotide polymorphisms (SNPs). To enable us in determining the key regulatory elements more efficiently for transcription factor binding studies we have performed in-silico predictions for potential regulatory regions

4.1.1 Chromosome 6q QTL

The QTL on chromosome 6q23 was identified by linkage analysis in an extended Asian family with β-thalassaemia. The 6q QTL exists as a set of single nucleotide polymorphisms (SNPs) spanning a region of ≈ 79kb that includes the intergenic region 5[i] of*MYB* and *HBS1L* genes.The SNPs are distributed in three linkage disequilibrium (LD) blocks referred to as *HBS1L –MYB* intergenic polymorphism *(HMIP)* blocks 1, 2, and 3. The 3 blocks completely account for the HbF variance attributed to the 6q QTL with SNPs in block 2 (24kb) showing the strongest effect and accounting for the majority of the variance. To direct us to the important spots for transcription factor binding studies in the chromosome 6q we seek aid from in *silico* experiments.

4.1.2 Transcription factor prediction tools

Transcription factor prediction tools are an important area of computational biology which enables us to delineate binding sites that are not reported previously. Computational methods to identify transcription factor binding sites were developed based on consensus sequences or Positional weight matrix (Quandt et al 1997). Positional weight matrix (PWM) is a model used to describe the sequence pattern of binding sites. As more transcription factors and their binding sites are experimentally identified, databases storing transcription regulation information were established, such as TRANSFAC, and TRRD. Consensus sequences were used to describe the

sequence patterns of cis-elements, and later PWMs were developed to describe the sequence motifs more precisely. Most of the prediction tools use the TRANSFAC database as the main matrix database for their predictions.

4.2Results

4.2.1 Sequence Analysis

Two individuals, one homozygous for the for the haplotype associated with high HbF across the 6q QTL (blocks 1, 2 and 3), designated as HPFH +/+ and another, homozygous for the absence of the SNPs associated with increased HbF levels, designated as HPFH -/- were selected for resequencing of a 30kb region that encompasses *HMIP* block 2 (*HMIP-2*) (Figure 4.1). Sequence analysis revealed that there were fifty five variants that differ between the HPFH +/+ and the HPFH -/- individuals. Table 4.1 is a list of the different SNPs between the two individuals. The HPFH +/+ individual was homozygous for all the SNPs while the HPFH -/- individual was homozygous for some but heterozygous for the other SNPs. The sequencing experiments were performed by colloborators at CNG (France).

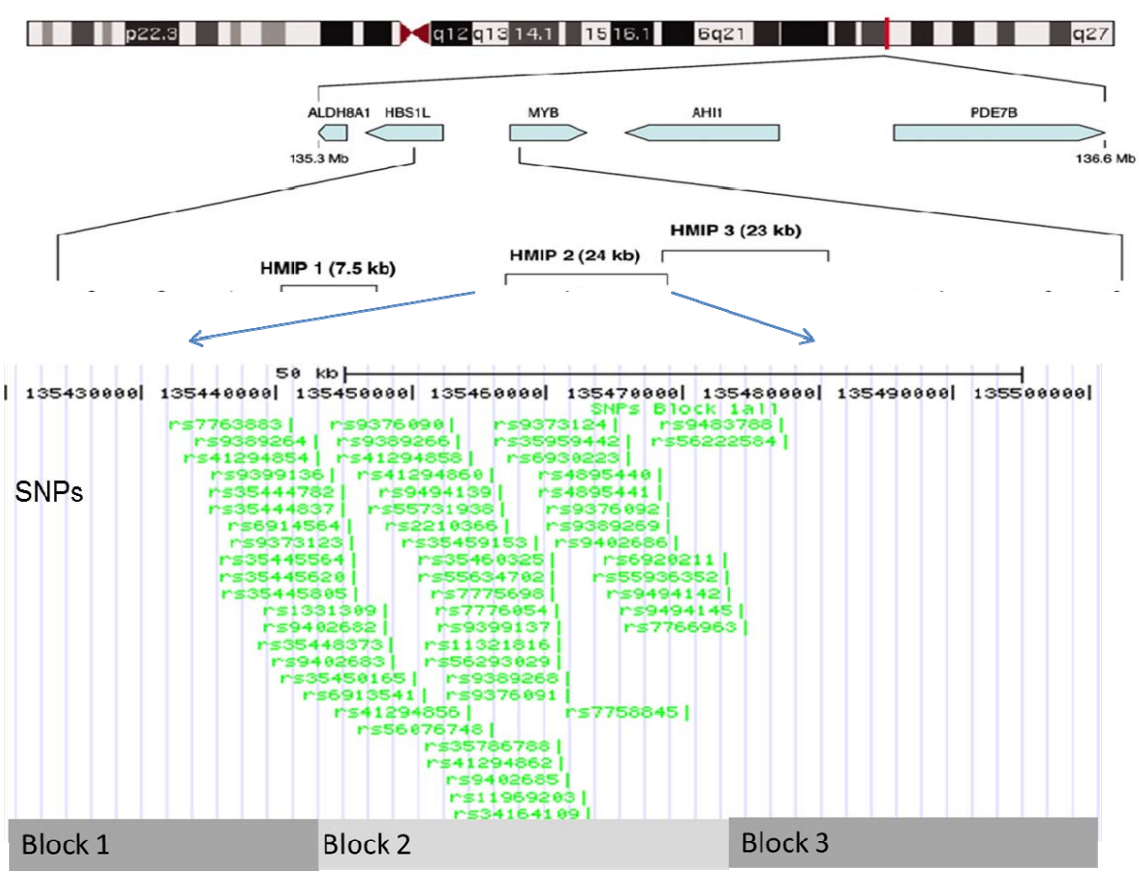


Figure 4.1:*HMIP-2* SNPs that were revealed by sequence analysis.

Position UCSC Genome Browser (March 2006)	HPFH -/-	HPFH +/+	Reference allele	SNP ID
***HMIP* block 1**				
135,440,936	CT	CC	C	rs7763883
135,442,805	CG	CC	C	rs9389264
135,443,064	CT	CC	C	rs41294854
135,444,032	TT	CC	T	rs9399136
135,444,782	CG	CC	C	c6_snp1
135,444,837	CG	GG	G	c6_snp2
135,445,317	CT	CC	C	rs6914564
***HMIP* block 2**				
135,445,512	CT	TT	T	rs9373123
135,445,564	CT	CC	C	CNG discovered
135,445,620	CT	TT	T	c6_snp3
135,445,805	AG	GG	G	c6_snp4
135,447,871	TT	GG	T	rs1331309
135,447,876	T-	GG	G	rs9402682
135,448,373	TT	CC	T	c6_snp5
135,448,491	AT	TT	T	rs9402683
135,450,165	GT	GG	G	c6_snp6

135,450,809	AG	AA	A	rs6913541
135,452,921	TT	CC	T	rs9376090
135,453,430	GT	GG	G	rs9389266
135,454,075	CT	CC	C	rs52090905
135,454,329	CT	TT	T	rs52090906
135,455,762	CG	CC	C	rs52090907
135,455,902	AG	GG	G	rs56076748
135,456,486	AG	AA	A	rs9494139
135,456,543	AG	GG	G	rs55731938
135,456,901	AA	GG	G	rs2210366
135,459,153	GT	GG	G	CNG discovered
135,460,328	CC	--	C	rs7775698 or rs66650371
135,460,330	TAG	---/---	TAG	rs67449035
135,460,609	AA	GG	A	rs7776054
135,460,711	TT	CC	T	rs9399137
135,460,731	CC	--	C	rs11321816
135,460,732	CC	AA	C	rs56293029
135,460,735	GG	AA	G	rs35786788
135,460,947	CT	TT	T	rs41294862
135,461,324	AA	GG	A	rs9389268
135,461,329	CC	TT	C	rs9376091
135,461,381	TT	CC	T	rs9402685
135,462,591	AA	TT	A	rs11969203
135,462,869	CC	TT	C	rs34164109
135,463,989	AA	TT	A	rs11759553
135,464,902	TT	CC	T	rs9373124
135,465,872	CC	GG	C	rs35959442
135,465,896	GG	TT	G	rs6930223
135,468,251	AA	TT	A	rs4895440
135,468,266	AA	GG	A	rs4895441
135,468,837	CC	AA	C	rs9376092
135,468,852	TT	CC	T	rs9389269
135,469,510	GG	AA	G	rs9402686
135,470,230	AA	CC	A	rs7758845
135,473,011	TT	CC	T	rs6920211
135,473,192	GG	AA	G	rs55936352
135,473,333	TT	CC	T	rs9494142
135,474,245	TT	CC	T	rs9494145
***HMIP* block 3**				
135,474,576	TT	CC	T	rs7766963
135,477,194	TT	CC	T	rs9483788
135,477,547	TT	AA	T	rs56222584

Table 4-1: *HMIP-2* SNPs that were revealed by sequence analysis

4.2.2 Transcription Factor Predictions at Chromosome 6q QTL

Transcription factor prediction tools (Matinspector and Consite) were used to predict transcription factor binding sites around the SNPs. Matinspector use the TRANSFAC, while Consite use Jaspar database. Initial predictions were performed using Matinspector with all SNPs. Consite was used as the second prediction tool to confirm binding of MYB , GATA-1 and TAL1 transcription factors around key SNPs.

Further comparison of predicted binding sites in the presence or absence of the implicated minor allele of the SNPs was performed to investigate if the SNPs alter binding of any transcription factors. 25 base pairs of sequences flanking the SNPs were included in the prediction tools and a high dissimilarity threshold was chosen to prevent irrelevant results.

4.2.2.1 Transcription Factor Prediction using Matinspector

Matinspector of genomatix site was used as the main tool for predicting transcription factor binding site. Matinspector is a software tool that utilizes a large library of matrix descriptions for transcription factor binding sites to locate matches in DNA sequences.The algorithm of Matinspector was used to calculate the matrix similarity with TRANSFAC matrices (TRANSFAC 6.0 public). The matrix similarity thus calculated will range from 0 to 1. A higher score indicated higher similarity between the sequence scanned and the sequence pattern represented in TRANSFAC matrix (Zhenget al2003).Mismatches in highly conserved positions of the matrix decrease the matrix similarity more than mismatches in less conserved regions. Matrix similarity threshold was set high for a robust and more stringent interpretation since setting the threshold low will lead to excessive amount of matches and higher number of false positives. Predictions were made for all transcription factor families in the matrix

(TRANSFAC 6.0).The same protocol was repeated with the reference sequence. The raw results were compared to see if any transcription factor binds differentially to the sequences in the presence and the absence of the minor allele of the SNPs.

Any major difference affecting matrix similarity was presented as demolishment of binding site enabling detection of differential binding. Table 4.2 shows some of the key transcription factors predicted in close proximity to the SNPs. Two regions that include key transcription factors are highlighted and will be discussed in detail (Figure 4.2).

Transcription Factor predictions from Matinspector		
Position	**SNP ID**	**Key Transcription Factors in the area**
***HMIP* block 1**		
135,440,936	rs7763883	PAX-2 binding sites Core promoter initiator elements
135,442,805	rs9389264	Homeodomain transcription factors Homeobox transcription factors
135,443,064	rs41294854	NKX homeodomain factors
135,444,032	rs9399136	Cellular and viral myb-like transcriptional regulators Krueppel like transcription factors
135,444,782	c6_snp1	SOX/SRY-sex/testis determining and related HMG box factors Brn POU domain factors
135,444,837	c6_snp2	PAX-2 binding sites
135,445,317	rs6914564	GATA binding factors Krueppel like transcription factors
***HMIP* block 2**		
135,445,512	rs9373123	p53 tumour suppressor Zinc finger proteins
135,445,564	CNG discovered	EGR/nerve growth factor induced protein C & related factors
135,445,620	c6_snp3	GATA binding factors MEF2, myocyte-specific enhancer binding factor

135,445,805	c6_snp4	Homeobox transcription factors DM domain-containing transcription factors
135,447,871	rs1331309	C-abl DNA binding sites
135,447,876	rs9402682	C-abl DNA binding sites
135,448,373	c6_snp5	Human acute myelogenous leukaemia factors(RUNX)
135,448,491	rs9402683	Hepatic Nuclear Factor 1 Homeobox transcription factors
135,450,165	c6_snp6	Nuclear factor of activated T-cells
135,450,809	rs6913541	Interferon regulatory factors Nuclear factor of activated T-cells
135,452,921	rs9376090	Nuclear factor 1
135,453,430	rs9389266	GATA binding factors Growth factor independence transcriptional repressor
135,454,075	rs52090905	Nuclear factor of activated T-cells Myeloid zinc finger 1 factors
135,455,902	rs56076748	MAF and AP1 related factors
135,456,486	rs9494139	Brn POU domain factors
135,456,543	rs55731938	Hepatic Nuclear Factor 1 Brn POU domain factors
135,456,901	rs2210366	Krueppel like transcription factors Heat shock factors
135,459,153	CNG discovered	Hepatic Nuclear Factor 1
-84 kb hot spot**		
135,460,328	*rs7775698 or rs66650371	Hepatic Nuclear Factor 1, **Twist subfamily of class B bHLH transcription factors, GATA binding factors, Human acute myelogenous leukaemia factors**
135,460,330	rs67449035	**GATA binding factors**
135,460,609	rs7776054	SOX/SRY-sex/testis determining and related HMG box factors POZ domain zinc finger expressed in B-Cells

135,460,711	rs9399137	**Cellular and viral myb-like transcriptional regulators,** Hepatic Nuclear Factor 1 Brn-5 POU domain factors
135,460,731	rs11321816	Cellular and viral myb-like transcriptional regulators
135,460,732	rs56293029	Cellular and viral myb-like transcriptional regulators
135,460,735	rs35786788	Cellular and viral myb-like transcriptional regulators
135,460,947	rs41294862	Hepatic Nuclear Factor 1 SOX/SRY-sex/testis determining and related HMG box factors
135,461,324	rs9389268	**Human acute myelogenous leukaemia factors** **Krueppel like transcription factors** **Cellular and viral myb-like transcriptional regulators**
135,461,329	rs9376091	Human acute myelogenous leukaemia factors
135,461,381	rs9402685	Krueppel-like C2H2 zinc finger factors hypermethylated in cancer
135,462,591	rs11969203	Vertebrate TATA binding protein factor
135,462,869	rs34164109	MAF and AP1 related factors
135,463,989	rs11759553	RXR heterodimer binding sites
135,464,902	rs9373124	Brn POU domain factors
135,465,872	rs35959442	Heat shock factors
135,465,896	rs6930223	Vertebrate TATA binding protein factor
135,468,251	rs4895440	Nuclear factor kappa B/c-related
135,468,266	rs4895441	NKX homeodomain factors
135,468,837	rs9376092	SOX/SRY-sex/testis determining and related HMG box factors ,Brn-5 POU domain factors
135,468,852	rs9389269	SOX/SRY-sex/testis determining and related HMG box factors
135,469,510	rs9402686	Heat shock factors
135,470,230	rs7758845	cAMP-responsive element binding proteins E-box binding factors
-71 kb hot spot**		
135,473,011	rs6920211	Nuclear receptor subfamily 2 factors Octamer binding protein
135,473,192	rs55936352	**Cellular and viral myb-like transcriptional**

		regulators
135,473,333	rs9494142	**GATA binding factors**
135,474,245	rs9494145	Ras-responsive element binding protein
135,474,576	rs7766963	PAX-2 binding sites
***HMIP* block 3**		
135,477,194	rs9483788	MEF2, myocyte-specific enhancer binding factor
135,477,547	rs56222584	GATA binding factors EVI1-myleoid transforming protein

Table4-2: Key transcription factors predicted in close proximity to the SNPs within the 3 HMIP blocks.

* = There are two SNPs located at chromosome 6 135,460,328. First one is a 3bp deletion rs66650371 (TAC/ ---) and the other one is rs7775698 (T/A). I have performed the predictions with rs66650371 (TAC/ ---).

** = -71kb and -84kb are two hot spots that are homologous to the key transcription factor binding regions in mouse demonstrated by Stadhouderset al. These regions also coincide with DNase I hypersensitive sites and GATA-1 sites previously detected by our group (Wahlberg et al 2009).

The tracks of high scoring SNPs prepared by Stephen Menzel was uploaded and viewed on the online UCSC genome browser (genome.ucsc.edu).

This provides a more detailed view of the region, with the flexibility of viewing SNP tracks, conserved transcription factor binding sites, hypersensitive regions, as well as many other tracks available, all in the same view. Tracks of the GATA-1 peaks at the 6q23 locus and surrounding area prepared by fellow students Karin Wahlberg and Kiran Jawaid were uploaded together for ease of orientation. Transcription factor predictions by Matinspector highlighted some SNPs which altered the binding of key transcription factors. Some key SNPs were located in two hotspots 71 kb and 84 kb away from *MYB* gene transcription start site and coincided with GATA-1 binding peaks detected by ChIP-chip experiments and intergenic transcripts as well as Dnase hypersensitive sites (Figure 4.2).

SNPs rs66650371, rs9393137 rs9389268 and rs9376091 were located around -84kb encompassing a region around 1kb (Figure 4.2 B). SNPs rs55936352 and rs9494142 were located around -71kb encompassing a region around 500bp (Figure 4.2 C). Figures 4.3 to 4.7 will focus on some of these important SNPs. The sequences of SNPs with flanking sequences are provided in the figures. Predicted sites for some important transcription factors are also highlighted. Tables 4.3 to 4.7 provide the prediction scores from the matinspector for the important factors around the mentioned SNPs.

Matinspector results suggested altered binding of MYB, KLF1 and AML (RUNX) around SNPs rs9389268 andrs9376091. Results also suggested altered binding of MYB around SNPs rs9393137 and rs55936352. GATA-1 binding motifs were predicted around SNPs rs66650371 and rs9494142 but matinspector prediction did not suggest disruption of GATA-1, TAL1 or AML binding motifs in the presence of the alternative alleles of these SNPs.

Figure4.2: Location of key SNPs

A: The key SNPs are located in two regions within block 2, around -84kb and -71kb away from the transcription start site of *MYB* gene, respectively.-71kb and -84kb are two hot spots that are homologous to the key transcription factor binding regions in mouse demonstrated by Stadhouders et al. In mouse LDB1 complexes which include GATA-1 and KLF1 transcription factors bind to these sites.

B: Gata-1 Peak 3 from previous ChIP-chip experiments. Peak 3 is 84 kb away from the MYB transcription start site (-84kb) Conserved GATA motifs, predicted transcription factor motifs and high scoring SNPs are found in close proximity.

C: Gata-1 Peak 4 from previous ChIP-chip experiments. Peak 4 is 71 kb away from the MYB transcription start site (-71kb) Conserved GATA motifs, predicted transcription factor motifs and high scoring SNPs are also found in close proximity.

4.2.2.2 Transcription Factor Predictions using Consite

Initial transcription factor predictions using matinspector highlighted some SNPs which altered the binding of transcription factors. Some of the critical SNPs were reassessed using another tool called Consite. The profile collection in Consite is drawn from the JASPAR database, another open-access, non-redundant collection of curated profiles. The objective of this prediction was to confirm MYB, GATA-1, TAL1 and KLF1 sites that were predicted by matinspector. Matrices belonging to transcription factors (GATA-1, TAL1, and MYB) were presented to Consite. However, KLF transcription factor family was not present in the matrices available in which case I had to depend on preliminary results from initial predictions by matinspector**.**

Table 4.8 shows transcription factor prediction results using the consite tool around SNPs rs9389268 and rs9376091. Initial predictions by matinspector suggested that MYB, KLF and RUNX binding is altered by the minor alleles of SNPs (Table 4.3).Consite also suggested that MYB binding is altered by the minor alleles of the SNPs rs9389268 and rs9376091 (Table 4.8) but failed to provide support for altered KLF1 and AML-1 binding. KLF family matrices were not available at Consite when predictions were performed and the difference in AML binding with major and minor alleles is not significant (0.830 compared to 0.810). Table 4.9 shows transcription factor prediction results using the consite tool around rs66650371. Both of the predictions tools suggested that none of the binding motifs for key transcription factors such as TAL-1 , GATA-1 and AML are altered significanlty by the 3bp deletion (rs66650371). Matinspector predicted that binding of Hepatic Nuclear Factor 1 (HNF1) to the 3bp deletion site (rs66650371) is altered but the result is not supported by the Consite prediction tool which does not predict a HNF1 binding site.

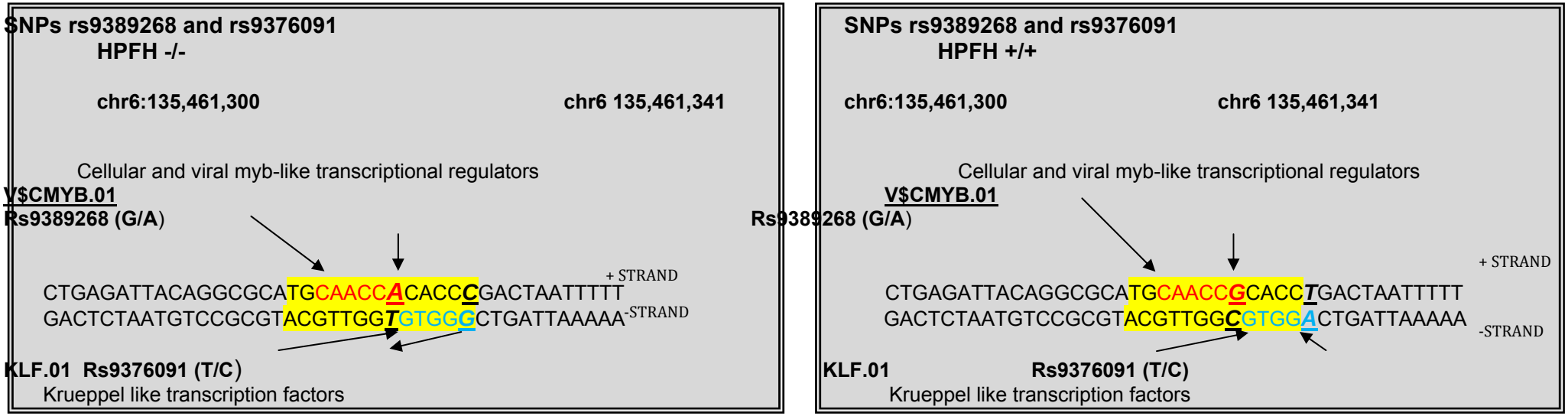


Figure4.3: Binding sites for key transcription factors around SNPs rs9389268 and rs9376091.
Letters in red represents the core binding sequence for MYB (CAAC-G) which is altered in the HPFH -/- genotype. Letters in blue represents the core binding sequence for KLF1 (GTGGG) which is altered in the HPFH +/+ genotype.

Matinspector Predictions for Transcription Factor Binding Sites				
Matrix Family	**Detailed Family Information**	**Core recognition sequence**	**Prediction Score(HPFH +/+) Rs9389268 =G Rs9376091 =T**	**Prediction Score(HPFH -/-) Rs9389268 =A Rs9376091 =C**
V$HAML	Human acute myelogenous leukaemia factors	GTGG	<0.700	0.903
V$KLFS	Krueppel like transcription factors	GTGGG	<0.700	0.960
V$BRNF	Brn POU domain factors	TAAT	<0.700	0.894
V$MYBL	Cellular and viral myb-like transcriptional regulators	CAAC-G	0.904	<0.700
V$HAND	Twist subfamily of class B bHLH transcription factors	CAGG	0.985	<0.700

Table4-3: Transcription Factors around the SNPs rs9389268 and rs9376091 and the binding site prediction scores for the HPFH +/+and HPFH -/- haplotypes.
Prediction scores are from matinspector where the minimum similarity threshold was set to 0.700 so transcription factors with a similarity less than this threshold was not shown in the raw data. Their prediction scores was presented as <0.700 because of this.

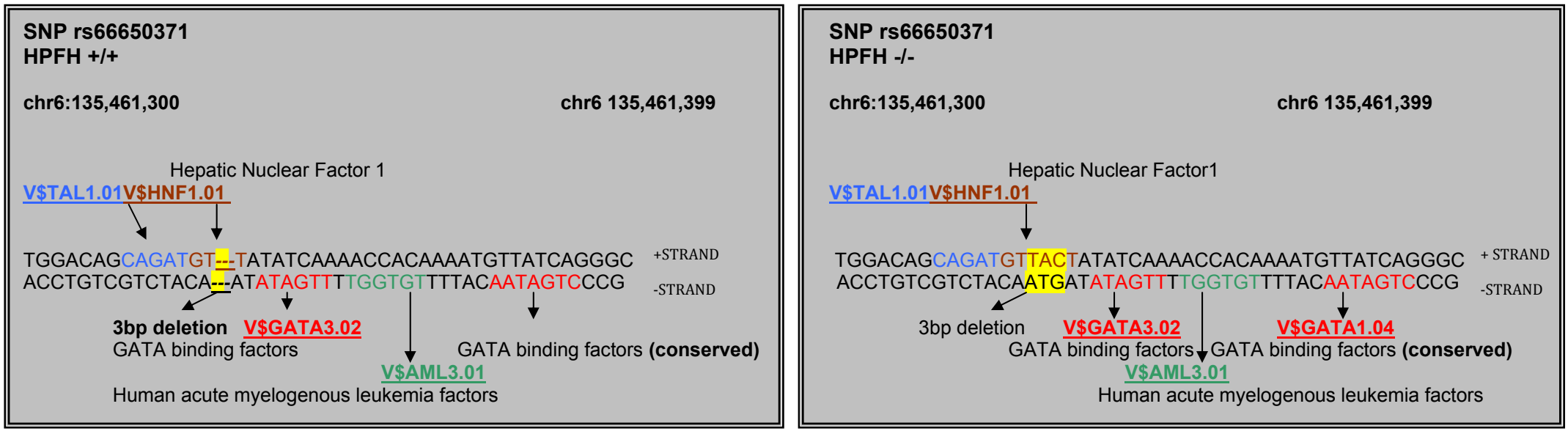


Figure 4.4: Binding sites for key transcription factors around 3bp deletion (rs66650371).
The core binding sequences for some transcription factors are highlighted and demolishment of core sequence for Hepatic Nuclear Factor 1(HNF1) in the HPFH +/+genotype can be observed.

Matinspector Predictions for Transcription Factor Binding Sites					
Matrix Family	**Detailed Family Information**	**Matrix**	**Core recognition sequence**	**Prediction Score (HPFH +/+)** **Rs66650371= ---**	**Prediction Score (HPFH -/-)** **Rs66650371= TAC**
V$HAND	Twist subfamily of class B bHLH transcription factors	V$TAL1BETAHEB.01	CAGAT	**0.941**	**0.946**
V$GATA	GATA binding factors	V$GATA3.02	AGAT	**0.910**	**0.910**
V$HNF1	Hepatic Nuclear Factor 1	V$HNF1.01	GTTA	**<0.700**	**0.812**
V$HAML	Human acute myelogenous leukaemia factors	V$AML3.01	GTGG	**0.914**	**0.914**
V$GATA	GATA binding factors	V$GATA1.04	GATA	**0.993**	**0.993**

Table 4-4: Transcription Factors around the SNP rs66650371 and the binding site prediction scores for the HPFH +/+and HPFH -/- haplotypes.
Prediction scores are from matinspector where the minimum similarity threshold was set to 0.700 so transcription factors with a similarity less than this threshold was not shown in the raw data. Their prediction scores was presented as <0.700 because of this.

Figure4.5: Binding sites for key transcription factors around SNP Rs9494142.
The core binding sequences for some transcription factors are highlighted and the position of SNP is close enough to the GATA-1 core motif to alter binding.

Matinspector Predictions for Transcription Factor Binding Sites				
Matrix Family	**Detailed Family Information**	**Matrix**	**Prediction Score (HPFH +/+) Rs9494142= C**	**Prediction Score (HPFH -/-) Rs9494142= T**
V$TALE	TALE homeodomain class recognizing TG motifs	V$TGIF.01	1.000	1.000
V$NKXH	NKX homeodomain factors	V$BAPX1.01	0.877	0.700>
V$RUSH	SWI/SNF related nucleophosphoproteins with a RING finger DNA binding motif	V$SMARCA3.02	0.986	0.968
V$GATA	GATA binding factors	V$GATA1.03	0.954	0.917
V$YY1F	Activator/repressor binding to transcription initiation site	V$YY1.02	0.955	0.955
V$SPZ1	Testis-specific bHLH-Zip transcription factors	V$SPZ1.01	0.945	0.945

Table4-5:Transcription Factors around the SNP Rs9494142 and the binding site prediction scores for the HPFH +/+and HPFH -/- haplotypes.
Prediction scores are from matinspector where the minimum similarity threshold was set to 0.700 so transcription factors with a similarity less than this threshold was not shown in the raw data. Their prediction scores was presented as <0.700 because of this. Transcription factor prediction suggested that GATA-1 binding was not altered significantly by Rs9494142.

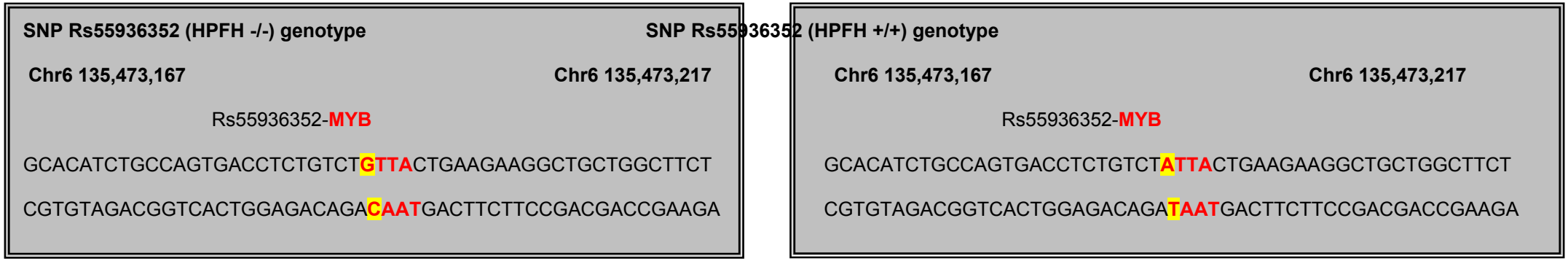
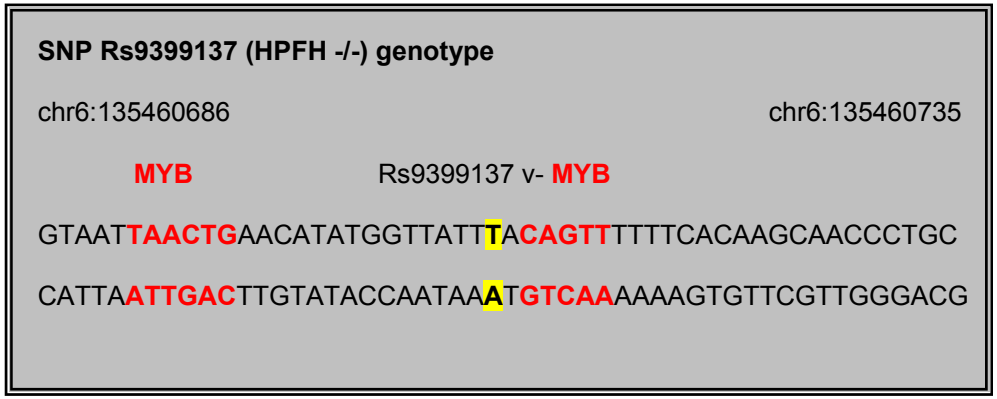


Figure4.6: Binding sites for key transcription factors around SNP Rs55936352.
The core binding sequences for MYB transcription factors are highlighted and demolishment of core sequence for MYB in the HPFH +/+ genotype can be observed.

Matinspector Predictions for Transcription Factor Binding Sites				
Matrix Family	**Detailed Family Information**	**Matrix**	**Prediction Score (HPFH +/+)** Rs55936352= **A**	**Prediction Score (HPFH -/-)** Rs55936352= **G**
V$RORA	v-ERB and RAR-related orphan receptor alpha	V$REV-ERBA.01	0.891	0.891
V$NR2F	Nuclear receptor subfamily 2 factors	V$HNF4.03	0.864	0.860
V$NBRE	NGFI-B response elements, nur subfamily of nuclear receptors	V$NBRE.01	0.867	0.867
V$RXRF	RXR heterodimer binding sites	V$RAR_RXR.01	0.830	0.830
V$CART	Cart-1 (cartilage homeoprotein 1)	V$PHOX2.01	0.891	0.700>
V$FAST	FAST-1 SMAD interacting proteins	V$FAST1.01	0.838	0.700>
V$PBXC	PBX1 – MEIS1 complexes	V$PBX1_MEIS1.02	0.772	0.700>
V$MYBL	Cellular and viral myb-like transcriptional regulators	V$CMYB.02	0.700>	0.961
V$PARF	PAR/bZIP family	V$HLF.01	0.700>	0.847

Table4-6: Transcription Factors around the SNP Rs55936352 and the binding site prediction scores for the HPFH +/+and HPFH -/- haplotypes.
Prediction scores are from matinspector where the minimum similarity threshold was set to 0.700 so transcription factors with a similarity less than this threshold was not shown in the raw data. Their prediction scores was presented as <0.700 because of this. Results suggest differential binding to MYB one of our key transcription factors.



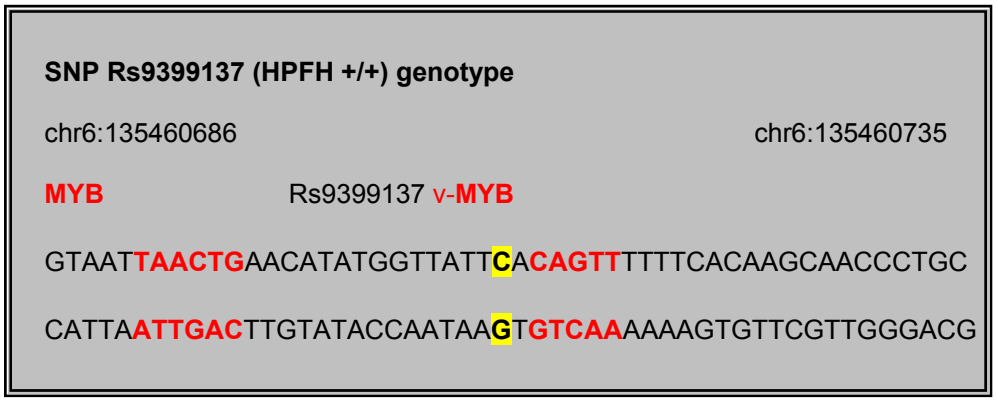


Figure4.7: Binding sites for key transcription factors around SNP rs9399137.
The core binding sequences for MYB transcription factors are highlighted.

Matinspector Predictions for Transcription Factor Binding Sites				
Matrix Family	**Detailed Family Information**	**Matrix**	**Prediction Score (HPFH +/+) Rs9393137=C**	**Prediction Score (HPFH -/-) Rs9393137=T**
V$HNF1	Hepatic Nuclear Factor 1	V$HNF1.04	0.912	0.912
V$MYBL	Cellular and viral myb-like transcriptional regulators	V$CMYB.02	0.969	0.969
V$MYBL	Cellular and viral myb-like transcriptional regulators	V$VMYB.01	0.700>	0.883
V$PLZF	C2H2 zinc finger protein PLZF	V$PLZF.01	0.700>	0.874

Table4-7: Transcription Factors around the SNP Rr9399137 and the binding site prediction scores for the HPFH +/+and HPFH -/- haplotypes.
Prediction scores are from matinspector where the minimum similarity threshold was set to 0.700 so transcription factors with a similarity less than this threshold was not shown in the raw data. Their prediction scores was presented as <0.700 because of this. Results suggest differential binding of MYB one of our key transcription factors.

Transcription Factor Predictions from Consite (Jaspar) for SNPs rs9389268 and rs9376091			
Matrix Family	**Detailed Family Information**	**Prediction Score (HPFH +/+) Rs9389268 =G Rs9376091 =T**	**Prediction Score (HPFH -/-) Rs9389268 =A Rs9376091 =C**
V$HAML	Human acute myelogenous leukaemia factors	0.810	0.830
V$MYBL	Cellular and viral myb-like transcriptional regulators	0.920	0.770

Table 4-8: Binding sites for transcription factors around the SNPs rs9389268 and rs9376091 and binding site prediction scores forthe HPFH +/+ and HPFH -/- haplotypes.

Transcription Factor Predictions from Consite (Jaspar) for SNP rs66650371 (3bp deletion)			
Matrix Family	**Detailed Family Information**	**Prediction Score (HPFH +/+) Rs66650371= ---**	**Prediction Score (HPFH -/-) Rs66650371= TAC**
V$HAND	Twist subfamily of class B bHLH transcription factors	**0.910**	**0.890**
V$GATA	GATA binding factors	**0.870**	**0.870**
V$HAML	Human acute myelogenous leukaemia factors	**1.000**	**1.000**
V$GATA	GATA binding factors	**0.840**	**0.840**

Table4-9: Binding sites for transcription factors around the SNP rs66650371 and the binding site prediction scores for the HPFH +/+ and HPFH -/- haplotypes.

4.3 Discussion

Resequencing of a 30kb region that encompasses block 2 SNPs at chromosome 6q QTL revealed 55 SNPs that differ between the high F and low F individuals. To direct functional investigation of all the key SNPs, we utilized *in-silico* tools. Computational predictions have been successfully utilized for suggesting potential regulatory regions for further experimental analysis enabling researchers to determine key regulatory elements more efficiently.

Matinspector of genomatix site was used as the main tool for predicting transcription factor binding sites. Predictions were performed to investigate presence of transcription factors binding motifs and to investigate if alternative alleles at these transcription factor binding motifs would affect binding. To reduce false positives, SNPs with 25 base pairs of flanking sequences were included in the prediction tools. This increases the stringency of detection since transcription-factor binding sites (TFBSs) are usually short (around 5-15 base-pairs); thus potential binding sites can occur very frequently. Matrix similarity threshold was also set to further increase the stringency since setting the threshold low will lead to excessive amount of matches.

Matinspector predicted 277 transcription factor binding sites around 55 SNPs, 64 of which were altered when the minor allele is present. Table 4.2 demonstrates some of the key transcription factors that are predicted in close proximity to the SNPs. Matinspector highlighted some SNPs which altered the binding of key transcription factors.

The key SNPs were located at regions located at 71 kb and 84 kb away from the *MYB* gene transcription start site. These sites were designated as hotspots because they coincided with GATA-1 binding peaks detected by previous ChIP-chip experiments,

intergenictranscripts, Dnase hypersensitive sites. In the mouse the homologous regions of these sites were bound by LDB1 transcription complexes which include key transcription factors like KLF1, GATA-1. These complexes have a key role in the interaction of intergenic region with the *MYB* gene promoter (Figure 4.2).
Consite prediction tool was used to validate initial predictions around SNPs rs9389268, rs9376091 and rs66650371 only with selected transcription factor families (MYB, GATA, AML, and TAL1). A total of 6 binding sites were predicted around 3 SNPs, 1 of which was altered when the minor allele is present.

There are two SNPs located at chromosome 6 135,460,328. The first is a deletion of 3bp designated as rs66650371 (TAC/ ---) and the other is SNP rs7775698 (T/A). I have performed the predictions with rs66650371 (TAC/ ---). Both of the predictions tools suggested presence of binding motifs for key transcription factors such as TAL-1 , GATA-1 and AML in close proximity of the 3bp deletion (rs66650371); but failed to suggest that these binding motifs are altered significantly since the deletion is not in the core motif of any transcription factor. Matinspector predicted demolishment of binding motif for Hepatic Nuclear Factor 1 (HNF1) but HNF1 is not a recognized transcription factor in gene regulation.
Presence of GATA-1 binding sites at -84kb was also suggested by Farrell et al but the study did not suggest disruption of key transcription factor binding motifs; instead they suggested that the region has enhancer activity and the enhancer potential is disrupted by the variants (Farrell et al 2011).Stadhouders et al demonstrated that they are LDB1 binding sites at -71kb and at -84 kb suggesting that LDB1 complexes including GATA-1 and TAL1 transcription factors are binding to these It is possible that the 3bp deletion rs66650371 at -84kb affect the spacing between these motifs leading to altered binding. Various studies have reported that spacing between TF binding motifs within enhancer is crucial for optimal binding of transcription factors which was investigated using EMSA (Chapter 5) and ChIP (Chapter 6) experiments.

Matinspector suggested the presence of MYB and KLF1 binding sites in close proximity to SNPs rs9389268 and rs9376091. The two SNPs are only 4 base pair from each other and genetic studies suggested that they are linked so I did not perform separate predictions for each SNP. Further predictions suggested that binding motifs of these transcription factors are altered in the presence of minor alleles. Differential binding of MYB was also confirmed by Consite prediction tool. MYB and KLF1 are crucial transcription factors and altered binding of any of them can potentially lead to differential regulation of flanking genes: *MYB* or *HBS1L.*Transcription factor binding to this region was further investigated using EMSA (Chapter 5) and ChIP (Chapter 6) experiments.Matinspector prediction tool also suggested the presence of GATA-1 binding motifs in close proximity of the SNP rs9494142 but failed to suggest that this binding motif is altered significantly when the alternative allele is present. Previous ChIP-chip experiments demonstrated presence of GATA-1 binding to this region and suggested altered binding of GATA-1 in individuals who are homozygous for the minor allele of SNP rs9494142. There is also another conserved GATA-1 site at ~100bp away from the SNP. Previous studies suggested that GATA-1 can bind as a dimer and thus this SNP could potentially affect the dimer binding to this region. Transcription factor binding to this region was further investigated using EMSA (Chapter 5) and ChIP (Chapter 6) experiments.Matinspector prediction tool suggested presence of two MYB binding motifs in close proximity of the SNP rs9399137. Predictions suggest that one of the binding motifs is altered significantly when the minor allele of SNP is present. Another MYB binding site was also predicted in close proximity of the SNP rs55936352. Predictions suggest that one of the binding motifs is altered significantly in the presence of the minor allele of the SNP. As discussed before, MYB is a crucial transcription factor so transcription factor binding to these regions were further investigated using ChIP (Chapter 6) experiments.

5 INVESTIGATION OF TRANSCRIPTION FACTOR BINDING AT THE 6Q23 LOCUS BY ELECTROPHORETIC MOBILITY SHIFT ASSAYS (EMSA)

5.1 Introduction

Prediction of transcription factor binding performed on all SNPs in 6q intergenic region showed that binding sites of some important transcription factors were altered by key SNPs in the region. We wanted to investigate binding of these predicted transcription factors in vitro using electrophoresis mobility shift assays (EMSA).

5.1.1 Chromosome 6q QTL

The QTL on chromosome 6q23 was identified by linkage analysis in an extended Asian family with β-thalassaemia through the proband who had very mild thalassaemia intermedia despite being homozygote for the β^0 thalassaemia with the complete absence of HbA. The extended family was large enough to perform linkage analysis, which mapped the HbF locus to a region spanning 1.5 Mb on chromosome 6q23. The 6q HbF QTL was validated in the large panel of European twins. Higher resolution genotyping refined the locus to a region of 79 kb between the *HBS1L* and *MYB* gene. The analysis revealed that this was a major QTL which accounted for 35% of total F-cell variance in the family (Thein et al 1994).

The 6q QTL exists as a set of single nucleotide polymorphisms (SNPs) spanning a region of ≈ 79kb that includes the intergenic region 5^i to *MYB and HBS1L and* 5^i part of the *HBS1L gene.*The SNPs are distributed in three linkage disequilibrium (LD) blocks referred to as *HBS1L –MYB* intergenic polymorphism *(HMIP)* blocks 1, 2, and 3. The 3 blocks completely account for the HbF variance attributed to the 6q QTL with SNPs in block 2 (24kb) showing the strongest effect and accounting for the majority of the variance.

Recent work by our group indicated the presence of regulatory elements in *HBS1L-MYB* intergenic region (*HMIP*) evidenced by GATA-1 binding coinciding with strong

histone acetylation, RNA polymerase II activity, and erythroid specific Dnase I hypersensitive sites (Wahlberg et al2009).

Genetic studies and resequencing of two key individuals from the family that led to discovery of 6q QTL revealed new SNPs in the *HBS1L-MYB* intergenic region. Prediction of transcription factor binding performed on all SNPs in 6q intergenic region showed that binding sites of some important transcription factors were altered by key SNPs in the region (Chapter 4). Figure 5.1 shows 4 key SNPs that were chosen for transcription factor binding investigation using electrophoresis mobility shift assays (EMSA). EMSAs using oligonucleotides encompassing these SNPs were carried out to demonstrate any differential transcription factor binding.

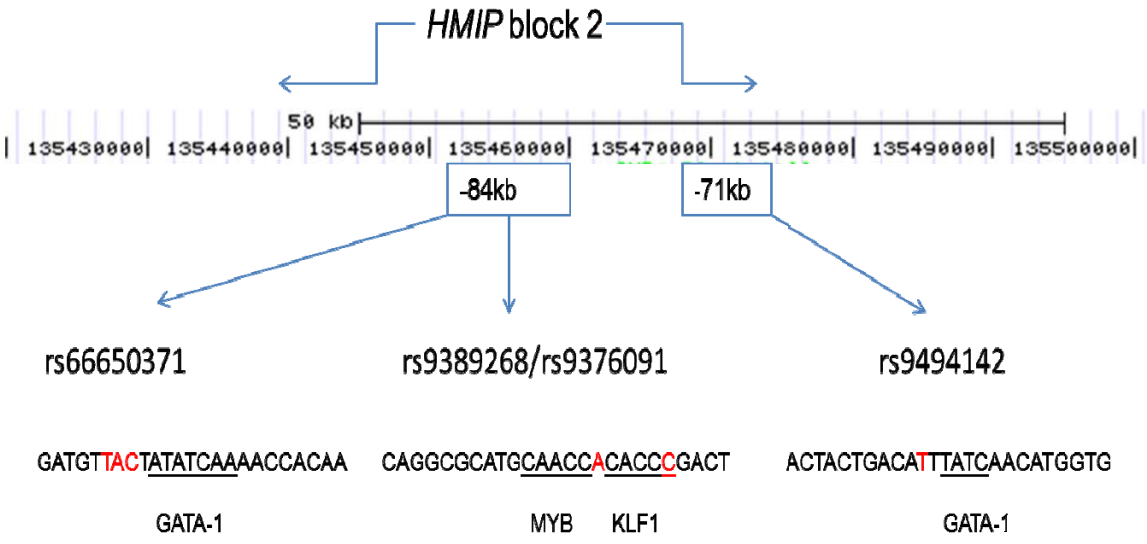


Figure 5.1: Transcription factor binding investigation around four SNPs, three (rs66650371, rs9389268 and rs9376091) in the hot spot at -84kb and one (rs9494142) in the hot spot at -71kb from the transcription start site of *MYB,* were performed using EMSAs. The main transcription factors of interest were: GATA-1, MYB and KLF1.The SNPs are shown in red and core binding motifs of GATA-1 (TATATCA), MYB (CAACCG) and KLF1 (CACCC) are underlined.

5.1.2 Electrophoretic Mobility Shift Assays (EMSA)

Electrophoresic mobility shift assay (EMSA) is a technique that is used for investigating protein-nucleic acid interactions. It is based on the observation that the electrophoretic mobility of a protein-nucleic acid complex is typically less than that of the free nucleic acid. The DNA is labelled to enable rapid detection and in its native state, it migrates quickly through the gel matrix. Protein binding slows the mobility of the protein-nucleic acid complexes that resolve as discrete bands. The method relies on the stability of the protein-DNA complexes when subject to non-denaturing polyacrylamide gel electrophoresis (Revzin et al 1989) (Figure 5.2).

Different detection methods such as radioactive labelling, fluorescence, chemiluminescence and immunohistochemical labeling can be used in EMSAs. Oligonucleotides labelled with biotin were used in our experiments to provide sensitive detection of the protein-nucleic acid complexes.

The mobility shift assay has a number of strengths. The basic technique is simple to perform, yet robust enough to accommodate a wide range of binding conditions.
A wide range of nucleic acid sizes and structures are compatible with the assay, proteins ranging in size from small oligopeptides to transcription factor complexes can give useful mobility shifts and the assay works well with both highly purified proteins and crude cell extracts. These properties account in a large part for the continuing popularity of the assay. EMSA provides an unbiased way to detect protein complexes binding to oligonucleotides but a candidate protein approach is still necessary in order to perform the EMSA experiments with the right crude nuclear extracts. Another major limitation of the EMSA experiments is the identification of proteins that are present in the protein complexes which requires additional steps that are not straight forward.

The electrophoretic “supershift” assay and assays that combine EMSA with western blotting or mass spectroscopy have been devised to allow identification of nucleic acid-associated proteins. Another alternative to detect proteins present in the protein complex is competing the proteins out with unlabeled oligonucleotides that are known to be bound by the proteins of interest (Hellman and Fried 2007).

We have performed the experiments using The LightShift Chemiluminescent EMSA Kit that uses a non-isotopic method to detect DNA-protein interactions. Biotin end-labelled DNA containing the binding site of interest is incubated with a nuclear extract or purified factor. The reaction is then subjected to gel electrophoresis on a native polyacrylamide gel and transferred to a nylon membrane. The biotin end-labelled DNA is detected using the Streptavidin-Horseradish Peroxidase Conjugate and the Chemiluminescent Substrate.

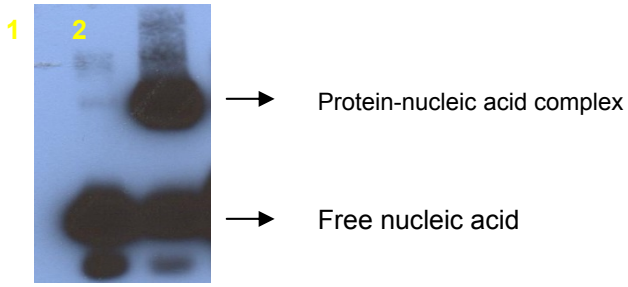


Figure 5.2: Electrophoretic Mobility Shift Assays (EMSA).
The difference in the electrophoretic mobility of free nucleic acid (lane1) and that bound to protein, i.e. protein-nucleic acid complex (lane 2) is demonstrated.

5.2 Results

5.2.1Optimization of EMSA

The Thermo Scientific LightShift Chemiluminescent EMSA Kit (Pierce) was used to optimize the EMSAs. The kit included nuclear extracts from EBNA cell lines where the control protein was highly expressed. Unlabelled and biotin labelled nucleotides were also provided. Binding reactions were prepared according to manufacturer's instructions. Reactions without cold competitors were incubated at room temperature for 20 minutes. Cold competitions were performed by addition of 200 fold excess unlabelled oligo, incubated 5 minutes at room temperature followed by a further 20 minute incubation at room temperature on addition of the biotin labelled oligonucleotide (Figure 5.3 and table 5.1).

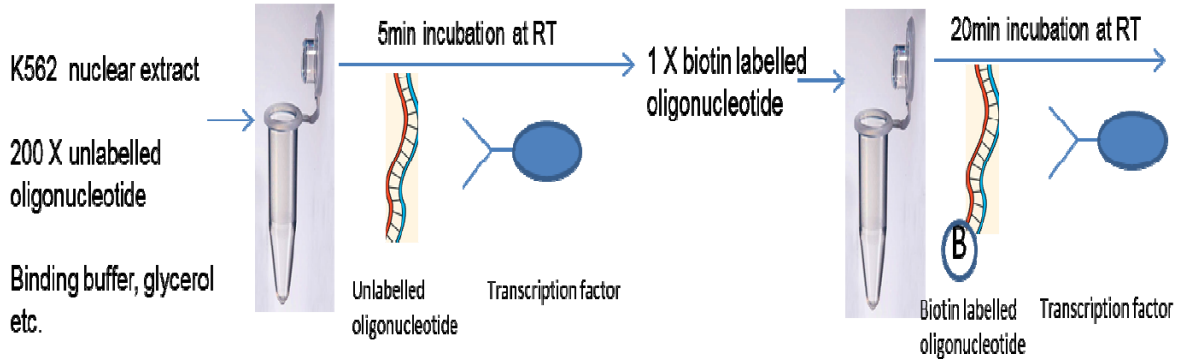


Figure 5.3: Cold competition EMSA assay

Results showed the presence of a strong protein-nucleic acid complex (Figure 5.4, lane 2) which is absent whenthe nuclear extract was not added (Figure 5.4, lane 1).The protein-nucleic acid complex is successfully competed out by the cold competitor (Figure 5.4, lane 3). Free oligonucleotides can be observed in all three lanes. The result shows that binding conditions are appropriate for the following EMSA experiments and that the lightshift kit is applicable to our experiments.

Binding reactions for Control EBNA System.				
Component	**Final Amount**	**Control Reactions**		
		No extract	Binding reaction	Cold competition
Ultrapure Water	----	12 µl	11 µl	9 µl
10X Binding Buffer (20148A)	1X	2 µl	2 µl	2 µl
50% Glycerol (20148F)	2.5%	1 µl	1 µl	1 µl
100 mM MgCl2 (20148I)	5 mM	1 µl	1 µl	1 µl
1 µg/µl Poly (dI•dC) (20148E)	50 ng/µl	1 µl	1 µl	1 µl
1% NP-40 (20148G)	0.05%	1 µl	1 µl	1 µl
Unlabeled EBNA DNA (20148C)	4 pmol	-----	-----	2 µl
EBNA Extract (20148D)	1 Unit	-----	1 µl	1 µl
Biotin–EBNA Control DNA (20148B)	20 fmol	2 µl	2 µl	2 µl
Total Volume	----	20 µl	20 µl	20 µl

Table5-1: Binding Reactions for Figure 5.4.
Three different binding reactions were prepared. The first reaction (negative control) was prepared without any extract. The second reaction was prepared with the extract provided with the kit. In the third reaction, unlabelled oligonucleotides were included as cold competitors to differentiate between specific and nonspecific protein-nucleic acid complexes.

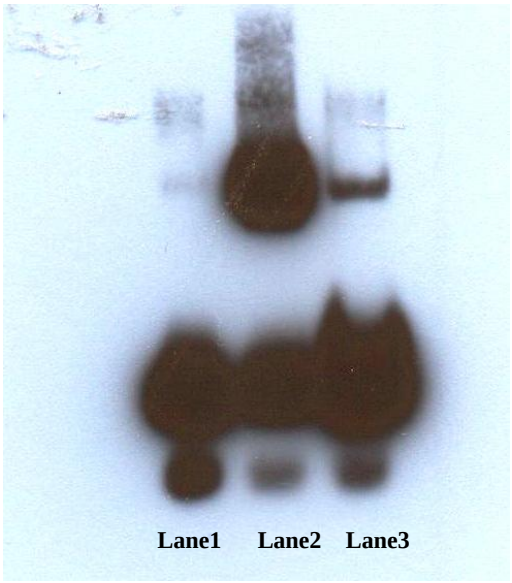


Figure5.4:Electrophoretic Mobility Shift Assays (EMSA).
The difference in the electrophoretic mobility of a protein-nucleic acid complex (lane 2) and free nucleic acid (lane1) is demonstrated. The competition of the complex by unlabelled oligonucleotide shows specificity of the binding (lane3).

5.2.2 Transcription factor binding at chromosome 6q

In-silico predictions performed on all SNPs in 6q intergenic region showed that the binding sites of some important transcription factors including MYB, KLF1 and GATA-1 were altered by the SNPs that had high association scores (Figure 5.1). EMSAs using oligonucleotides encompassing these SNPs were carried out to demonstrate any differential transcription factor binding.

Nuclear extracts used in the experiments were derived from K562 cells. Before setting up the EMSAs, western blots were performed to confirm presence of the transcription factor protein under investigation.

5.2.2.1 Transcription factor binding around rs9389268 (A/G)

Initial experiments were performed with oligos designed for the site encompassing SNP rs9389268 (A/G) and included 10 bases flanking the MYB recognition site (CAACCG). This oligo unfortunately also encompassed another SNP Rs9376091 (T/C) and KLF binding site (GTGGG) (Figure 5.5). The "old-g" oligonucleotide encompassed a binding site for MYB but not for the KLF; while "old-a" oligonucleotide encompassed a binding site for KLF1 but not for MYB (See Chapter 4 for details).

old-g oligo

CAGGCGCATGCAACCGCACCTGACT
MYB KLF1

old-a oligo

CAGGCGCATGCAACCACACCCGACT
MYB KLF1

Figure 5.5: EMSA oligonucleotides encompassing SNPs rs9389268 (A>G) highlighted in green and rs9376091(C>T) highlighted in red.

Commercial MYB positive (TACAGGCATAACGGTTCCGTAGTGA) and MYB negative oligos (TACAGGCATATCGGTTCCGTAGTGA) were also used in the experiments.

KLF1 positive and negative oligos were designed to encompass a known KLF1 binding site at the beta globin gene promoter.

Experiments were performed under the same binding conditions from initial optimization experiments. Table 5.2 shows each binding reaction for Figure 5.6.

	1	2	3	4	5	6	7	8	9
	old-g oligonucleotide			**old-a oligonucleotide**			**Myb+ oligonucleotide**		
Extract	**-**	**K562**	**K562**	**-**	**K562**	**K562**	**-**	**K562**	**K562**
Oligonucleotide	**old-g**	**old-g**	**old-g**	**old-a**	**old-a**	**old-a**	**Myb+**	**Myb+**	**Myb+**
Competitor	**-**	**-**	**old-g**	**-**	**-**	**old-a**	**-**	**-**	**Myb+**

Table 5-2: Binding Reactions for figure 5.6.
Three different binding reactions were prepared for each oligonucleotide i.e. (old-g, old-a, Myb+).

old-g oligo

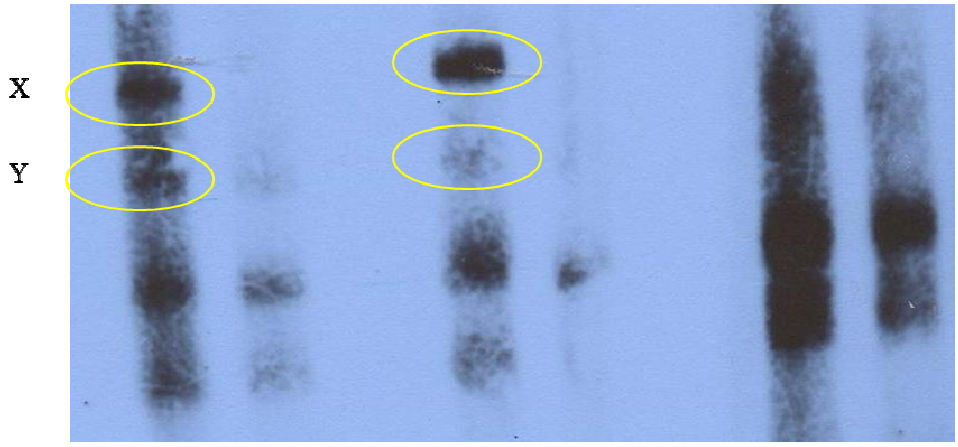
CAGGCGCATGCAACCGCACCTGACT

MYB KLF1

old-a oligo

CAGGCGCATGCAACCACACCCGACT

MYB KLF1



Lane	1	2	3	4	5	6	7	8	9
Extract	-	K562	K562	-	K562	K562	-	K562	K562
Oligonucleotide	old-g	old-g	old-g	old-a	old-a	old-a	Myb+	Myb+	Myb+
Competitor	-	-	old-g	-	-	old-a	-	-	Myb+

Figure5.6: Binding pattern of old-g, old-a, myb+ oligos with K562 nuclear extracts under binding conditions from table 5.2.

Lane 1 and Lane 4 are negative controls with no nuclear extract; thus no protein-nucleic acid complex is observed. Two specific protein-nucleic acid complexes are observed with both old-g and old-a oligonucleotides (X and Y) (Lane 2 and Lane 5, respectively) which are competed out with the presence of excess unlabelled oligonucleotide (Lane 3 and Lane 6, respectively). No specific protein-nucleic acid complex is observed with Myb+ oligonucleotide (Lane 8).

This EMSA results demonstrates the presence of specific protein-nucleic acid complexes with old-g and old-a oligonucleotides (Lane 2 and Lane 5, respectively). The bands in yellow circles seem to be specific and competed out by excess unlabelled oligonucleotide. The top band(X) is weaker with old-g oligo while the band below (Y) is stronger in old-g oligo. No specific binding was observed with commercial Myb+ oligonucleotide, as evidenced by the lack of bands after being competed out by excess unlabelled competitor. This result suggests potential differential binding between old-g and old-a oligonucleotides but fails to demonstrate that differentially bound complexes

are related to MYB transcription factor. The lack of specific protein-nucleic acid complex with Myb+ oligonucleotide suggests binding conditions were not ideal. We tested three alternative binding conditions as the next step to select the best condition for future experiments (Table 5.3).

Component	**Binding condition 1**	**Binding condition 2**	**Binding condition 3**
Ultrapure Water	9 µl	9 µl	8µl
10X Binding Buffer	2 µl	2 µl	2 µl
50% Glycerol	1 µl	1 µl	1 µl
MgCl2	1 µl	-	1 µl
Poly (dI•dC))	1 µl	1 µl	1 µl
1% NP-40	1 µl	-	1 µl
KCl	-	1 µl	-
EDTA	-	1 µl	1 µl
K562 nuclear extract	3 µl	3 µl	3 µl
Oligonucleotide	2 µl	2 µl	2 µl
Total Volume	20 µl	20 µl	20 µl

Table5-3: Three different binding conditions were prepared by alterations of KCL, EDTA, MgCl2, 1% NP-40 presence in the binding reaction.

	1	2	3	4	5	6	7	8	9	10
Extract	**K562**	**K562**	**K562**	**K562**	**K562**	**K562**	**K562**	**K562**	**K562**	**-**
Oligo	**old-g**	**old-g**	**old-g**	**old-a**	**old-a**	**old-a**	**Myb+**	**Myb+**	**Myb+**	**Myb+**
Competitor	**-**	**-**	**-**	**-**	**-**	**-**	**-**	**-**	**-**	**-**
Binding condition	**1**	**2**	**3**	**1**	**2**	**3**	**1**	**2**	**3**	**1**

Table 5-4: Binding Reactions for figure 5.7.

Three different binding reactions were prepared for each oligonucleotide: old-g, old-a and Myb+ with variable binding conditions as indicated in table 5.3 (Binding condition 1, 2, 3).

old-g oligo

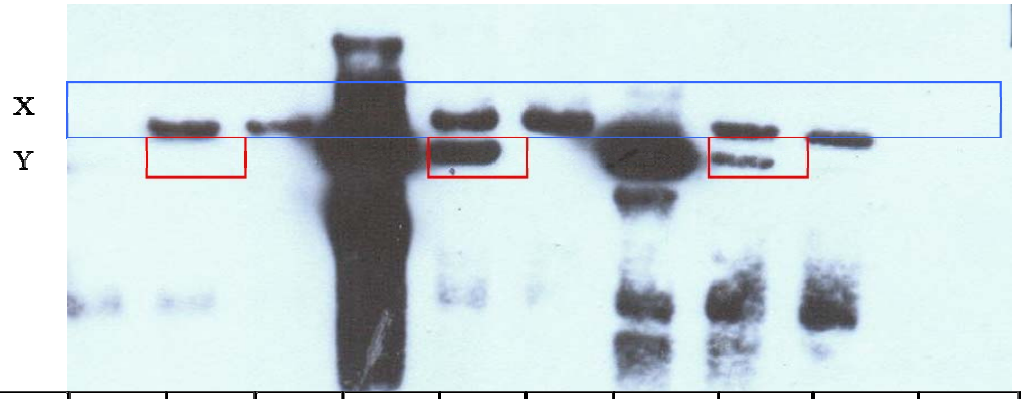
CAGGCGCATGCAACCGCACCTGACT

MYB KLF1

old-a oligo

CAGGCGCATGCAACCACACCCGACT

MYB KLF1



	1	2	3	4	5	6	7	8	9	10
Extract	**K562**	**K562**	**K562**	**K562**	**K562**	**K562**	**K562**	**K562**	**K562**	**-**
Oligo	**old-g**	**old-g**	**old-g**	**old-a**	**old-a**	**old-a**	**Myb+**	**Myb+**	**Myb+**	**Myb+**
Competitor	**-**	**-**	**-**	**-**	**-**	**-**	**-**	**-**	**-**	**-**
Binding condition	**1**	**2**	**3**	**1**	**2**	**3**	**1**	**2**	**3**	**1**

Figure5.7: Binding pattern of old-g, old-a, Myb+ oligos with K562 nuclear extracts and variable binding conditions.

Protein-nucleic acid complex (X) is observed with old-a, old-g and Myb+ oligonucleotides (lane 2 to lane 9) with all three binding conditions. Protein-nucleic acid complex (Y) is observed with old-a and Myb+ oligonucleotides under binding condition 2 only.

The titration exercise illustrated the importance of finding appropriate and optimal conditions in order to demonstrate the mobility shift for the different oligonucleotides. Each oligonucleotide had different binding patterns under different conditions. The oligonucleotide share a similar binding pattern when the experiments were performed under binding condition 3 (Lanes 3, 6 and 9) while different binding patterns were observed under alternative binding conditions (Lanes 2, 5 and 8).

The presence of a strong band (X) common to old-g, old-a and Myb+ oligonucleotides suggests that the old-a old-g and Myb+ oligonucleotides bind to the MYB-related transcription factor.

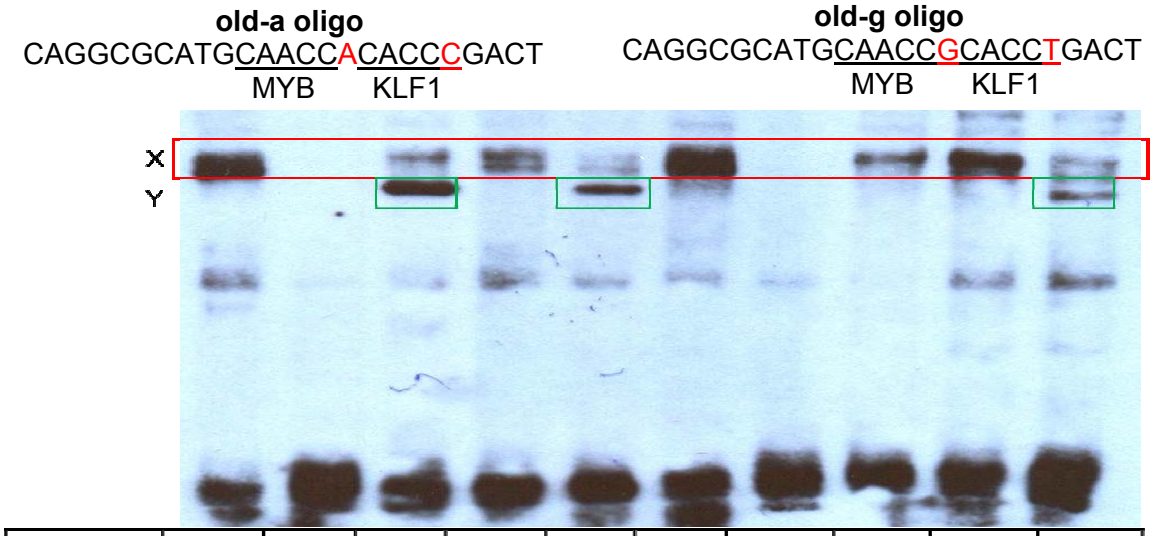
A second band (Y) can be observed with the old-a and Myb+ oligonucleotides under the binding condition 2; however no band is observed with the old-g oligonucleotide, suggesting that an alternative MYB related complex can be present. This band is also absent under other binding conditions which suggest that the binding is not robust. I have decided to perform experiments using binding condition 2 in order to investigate any potential differential binding with the old-a and old-g oligonucleotides.

5.2.2.1.1 Cold Competition EMSA

The initial EMSA experiments suggested a subtle differential binding between old-a and old-g oligonucleotides (Figure 5.6 and 5.7). I then performed cold competition experiments to investigate if the binding was specific to MYB or KLF1. The cold competition experiments were performed using excess unlabelled oligonucleotides that encompassed consensus binding sequences of MYB and KLF1. Each cold competitor was added to the binding reactions separately to validate binding specificity. Competition was repeated with oligonucleotides with mutated consensus sequences of MYB and KLF1. Table 5.5 shows the binding reactions that were prepared.

	1	2	3	4	5	6	7	8	9	10
Extract	**K562**	**K562**	**K562**	**K562**	**K562**	**K562**	**K562**	**K562**	**K562**	**K562**
Oligo	**old-a**	**old-a**	**old-a**	**old-a**	**old-a**	**old-g**	**old-g**	**old-g**	**old-g**	**old-g**
Competitor	**-**	**old-a**	**Myb+**	**Myb – mutant**	**KLF+**	**-**	**old-g**	**Myb+**	**Myb – mutant**	**KLF+**
Binding condition	**2**	**2**	**2**	**2**	**2**	**2**	**2**	**2**	**2**	**2**

Table 5-5: Binding Reactions for figure 5.8.
Cold competition of old-a and old-g oligonucleotides with Myb + (consensus MYB site), Myb – (mutated consensus site), KLF+ (consensus KLF1 site) were performed. All reactions were performed under binding condition 2 using 200 fold excess of unlabelled oligonucleotide which was added to the reaction 5 minutes before adding the labelled oligonucleotides.



Lane	1	2	3	4	5	6	7	8	9	10
Extract	K562	K562	K562	K562	K562	K562	K562	K562	K562	K562
Oligo	old-a	old-a	old-a	old-a	old-a	old-g	old-g	old-g	old-g	old-g
Competitor	-	old-a	Myb+	Myb – mutant	KLF+	-	old-g	Myb+	Myb – mutant	KLF+

Figure5.8: Cold competition patterns of old-a, old-g oligonucleotides with Myb +,Myb –, KLF+, old-a and old-g unlabelled oligonucleotides.
Specific protein-nucleic acid complex (X) is observed with old-a and old-g oligonucleotides (lane1 and lane 6, respectively) that are competed out by addition of excess self-competitor (lane 2 and lane 7, respectively). Addition of Myb+ and KLF +cold competitors with labelled old-a oligonucleotide almost competed out band X and led to the presence of new complex (Y) (Lane 3 and 5). Addition of Myb-(mutant) oligonucleotide cannot compete out band X and no new complex is observed (Lane 4). Addition of Myb+ and KLF+ unlabelled oligonucleotides with labelled old-g oligonucleotide partially compete out band X (Lane 8 and 10). Addition of KLF+ also led to the presence of new complex (Y) (Lane 10) that was not observed with the addition of Myb+ (Lane 8).

The result demonstrates the presence of MYB and KLF1 specific binding of the old-a and old-g oligonucleotides. Furthermore, the similar binding patterns of old-aoligonucleotide in the cold competition with Myb+ and KLF+ oligonucleotides suggeststhat the complex contained both of the transcription factors.

The lack of complete competition of complex X by Myb- (mutant) oligonucleotide with both old-a and old-g oligonucleotides provides further support that the complex observed is MYB specific.

The only difference observed in mobility shifts was the lack of formation of complex Y with Myb+ competition with old-g oligonucleotide. The complex binding patterns makes it hard to interpret any difference between the two oligonucleotide thus the effect of SNPs rs9389268 (A/G) and rs9376091 (T/C) should be validated by alternative methods. The results led us to conclude that the region encompasses motifs forming a complex pattern of MYB binding.

5.2.2.2 Cross Competition EMSA

The EMSA experiments suggested that the MYB and KLF1 bind to oligonucleotides encompassing SNPs rs9389268 and rs9376091; however no obvious differential binding was observed due to the complex binding patterns. In order to detect subtle differences between old-g and old-a oligonucleotides, cross competition experiments were designed. Each oligonucleotide was competed by its own cold competitor and by unlabelled oligonucleotide of the alternative competitor.

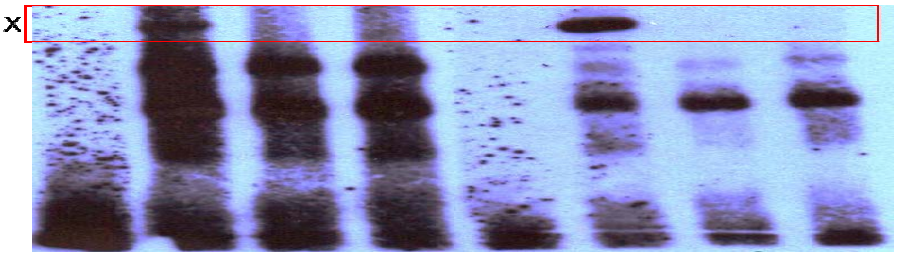
Initial experiments were performed with 200 fold excess of unlabelled competitor for each oligonucleotide to investigate any difference in the competition patterns. Table 5.6 shows the reactions prepared. The results failed to suggest any differential binding between the different oligonucleotides, since cross competition of each oligonucleotide was as efficient as cold competitions using the own sequence oligonucleotide (Figure 5.9).

	1	2	3	4	5	6	7	8
Extract	**-**	**K562**	**K562**	**K562**	**-**	**K562**	**K562**	**K562**
Oligo	**old-g**	**old-g**	**old-g**	**old-g**	**old-a**	**old-a**	**old-a**	**old-a**
Competitor	**-**	**-**	**Old-g 200x**	**Old-a 200X**	**-**	**-**	**Old-a 200X**	**Old-g 200X**
Binding condition	**2**	**2**	**2**	**2**	**2**	**2**	**2**	**2**

Table5-6: Binding Reactions for figure 5.9.
Cross competition of old-a and old-g oligonucleotides with unlabelled old-g and old-g oligonucleotides were performed. All reactions were performed under binding condition 2 using 200 fold excess of unlabelled oligonucleotide which was added to reaction 5 minutes before adding the labelled oligonucleotides.

old-g oligo
CAGGCGCATGCAACCGCACCTGACT
MYB KLF1

old-a oligo
CAGGCGCATGCAACCACACCCGACT
MYB KLF1



Lane	1	2	3	4	5	6	7	8
Extract	-	K562	K562	K562	-	K562	K562	K562
Oligo	old-g	old-g	old-g	old-g	old-a	old-a	old-a	old-a
Competitor	-	-	Old-g 200x	Old-a 200X	-	-	Old-a 200X	Old-g 200X

Figure5.9: Cross competition patterns of old-g, old-a oligonucleotides with old-g and old-a unlabelled oligonucleotides.
Free oligonucleotides can be observed with reactions without nuclear extract (Lane 1 and Lane 5). Specific protein-nucleic acid complex (X) is observed with old-a and old-g oligonucleotides (lane 2 and lane 6, respectively) that are competed out by addition of excess self-competitor (lane 3 and lane 7, respectively) as well as the other competitor (Lane 4 and 8, respectively).

5.2.2.3 Transcription factor binding around rs Rs9494142 (T/C)

Previous ChIP-Chip experiments demonstrated the presence of five GATA-1 binding peaks at the block 2 of chromosome 6q, three of which were around the conserved GATA-1 binding motifs. *In-silico* tools predicted another GATA-1 motif in close proximity to the conserved GATA-1 motifs. There were no SNPs around the conserved GATA-1 binding site but there was one SNP rs9494142 (T/C) that was encompassed by the predicted GATA-1 binding motif. EMSA experiments were carried out by using oligonucleotides encompassing the predicted GATA-1 binding site in order to investigate if binding is altered by SNP rs9494142 (T/C) (Figure 5.10).

"t" oligo	"c" oligo
ACTACTGACATTTATCAACATGGTG	ACTACTGACACTTATCAACATGGTG
GATA-1	GATA-1

Figure 5.10: EMSA oligonucleotides encompassing SNP rs9494142

Similarly, experiments were performed initially under variable binding conditions to select the most appropriate and optimum binding condition for the oligonucleotides of interest.

Table 5.7 demonstrates alternative binding conditions that were tested. Table 5.8 shows the competition of the reaction mix. Results demonstrated that there is no significant difference in the binding patterns of the oligonucleotides with variable binding conditions (Figure 5.11) and experiments were performed with binding condition 3.

Component	Binding condition 1	Binding condition 2	Binding condition 3
Ultrapure Water	9 µl	9 µl	8µl
10X Binding Buffer	2 µl	2 µl	2 µl
50% Glycerol	1 µl	1 µl	1 µl
MgCl2	1 µl	-	1 µl
Poly (dI•dC))	1 µl	1 µl	1 µl
1% NP-40	1 µl	-	1 µl
KCl	-	1 µl	-
EDTA	-	1 µl	1 µl
K562 nuclear extract	3 µl	3 µl	3 µl
Oligonucleotide	2 µl	2 µl	2 µl
Total Volume	20 µl	20 µl	20 µl

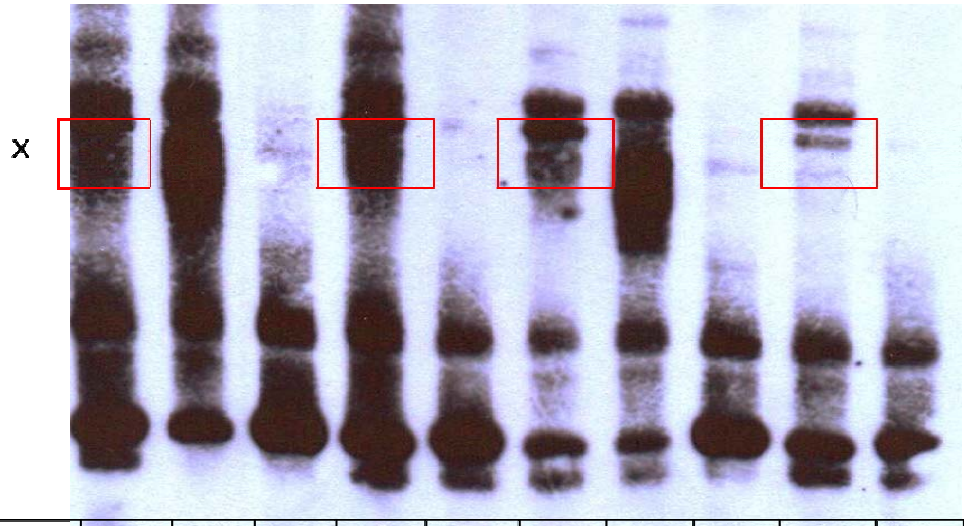
Table5-7: Three different binding conditions were prepared by alterations of KCL, EDTA, MgCl2, 1% NP-40 presence in the binding reaction.

	1	2	3	4	5	6	7	8	9	10
Extract	**K562**	**K562**	**K562**	**K562**	**K562**	**K562**	**K562**	**K562**	**K562**	**K562**
Oligo	**t**	**t**	**t**	**t**	**t**	**c**	**c**	**c**	**c**	**c**
Competitor	**-**	**-**	**t**	**-**	**t**	**-**	**-**	**c**	**-**	**c**
Binding condition	**1**	**2**	**2**	**3**	**3**	**1**	**2**	**2**	**3**	**3**

Table5-8: Binding Reactions for figure 5.11.
Three different binding reactions were prepared for each oligonucleotide encompassing the alternative allele T or C, referred to as “t” and “c” oligo under variable binding conditions i.e. (Binding condition 1, 2, 3).

“t” oligo
ACTACTGACATTTATCAACATGGTG
GATA-1

“c” oligo
ACTACTGACACTTATCAACATGGTG
GATA-1



Lane	1	2	3	4	5	6	7	8	9	10
Extract	K562	K562	K562	K562	K562	K562	K562	K562	K562	K562
Oligo	t	t	t	t	t	C	c	c	c	c
Competitor	-	-	t	-	t	-	-	c	-	c
Binding condition	1	2	2	3	3	1	2	2	3	3

Figure5.11: Binding pattern of t and c oligonucleotides with K562 nuclear extracts under variable binding conditions.
Specific protein-nucleic acid complex (X) is observed with both t and c oligonucleotides (lane 1,2,4,6,7,9) under all three binding conditions; the binding is competed out by unlabelled oligonucleotides (Lane 3,5,8,10).

5.2.2.3.1 Cold Competition EMSA

To characterize the complex (X) from Figure 5.11, cold competition experiments were performed. As bioinformatics suggested the presence of GATA-1 binding motifs, competition assays were performed using the commercial GATA-1 consensus sequences and the mutant GATA-1 sequences as competitors in order to investigate if the demonstrated protein-nucleic acid complex is GATA-1 related. Table 5.9 shows the comparison of the binding reaction mix.

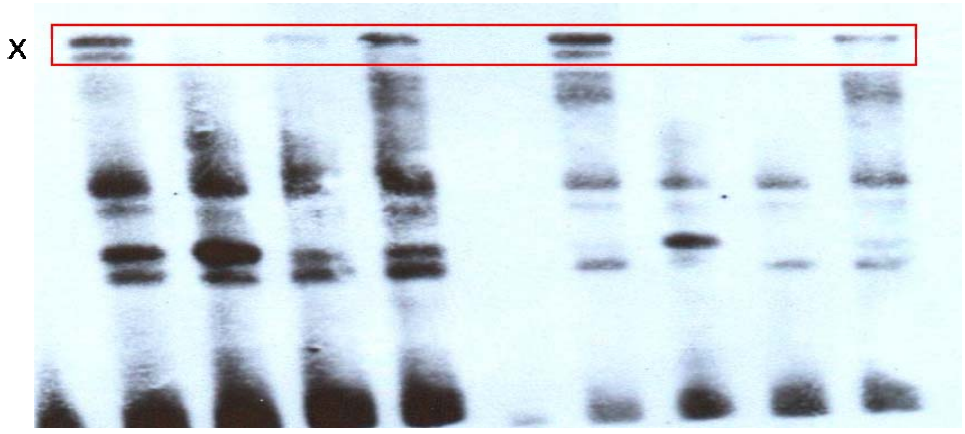
	1	2	3	4	5	6	7	8	9	10
Extract	**-**	**K562**	**K562**	**K562**	**K562**	**-**	**K562**	**K562**	**K562**	**K562**
Oligo	**t**	**t**	**t**	**t**	**t**	**c**	**c**	**c**	**c**	**c**
Competitor	**-**	**-**	**t**	**GATA-1+**	**GATA-1- Mutant**	**-**	**-**	**c**	**GATA-1+**	**GATA-1- Mutant**
Binding Condition	**3**	**3**	**3**	**3**	**3**	**3**	**3**	**3**	**3**	**3**

Table5-9: Binding Reactions for figure 5.12. Cold competition of t and c oligonucleotides with GATA-1 + (consensus GATA-1 site), GATA-1 – (mutated consensus site) were performed.

All reactions were performed under binding condition 3 with the addition of 200 fold excess of unlabelled oligonucleotide that was added to reaction 5 minutes before addition of the labelled oligonucleotides.

"t" oligo
ACTACTGACATTTATCAACATGGTG
GATA-1

"c" oligo
ACTACTGACACTTATCAACATGGTG
GATA-1



Lane	1	2	3	4	5	6	7	8	9	10
Extract	-	K562	K562	K562	K562	-	K562	K562	K562	K562
Oligo	t	t	t	t	t	c	c	c	c	c
Competitor	-	-	t	GATA-1+	GATA-1- Mutant	-	-	c	GATA-1+	GATA-1- Mutant

Figure5.12: Cold competition patterns of t and c oligonucleotides with GATA-1+, GATA-1- , t and c unlabelled oligonucleotides.

Specific protein-nucleic acid complex (X) is observed using t and c oligonucleotides (lane 2 and lane 7) that are competed out by addition of excess self-competitor (lane 3 and lane 8) as well as GATA-1+ competitor.Complete competition is not observed with the addition of GATA-1-(mutant) cold competitor (lanes 5 and 10, respectively).
The EMSA results demonstrated GATA related binding to “t” and “c” oligonucleotides, which were designed to encompass the SNP rs9494142 and a predicted GATA-1 site. Lane 1 and 6 are the negative controls without any nuclear extracts. The reactions were performed using nuclear extracts prepared from K562 cells where the expression of GATA-1 transcription factor is validated by western blots (lanes 2-5 and 7-10). Lanes 2 and 7 of figure 5.12 show a similar binding pattern of “t” and “c” oligonucleotides, respectively to K562 nuclear extracts and suggest similar binding pattern. The binding appears to be specific, as it is competed out by excess specific cold competitor (lanes 3 and 8, respectively). The complex appears to be GATA-1 related, as it is competed out using oligonucleotides encompassing the conserved GATA motifs but less so with mutated GATA motifs as cold competitors (lanes 4 and 8, respectively). The results failed to suggest that SNP rs9494142 alters GATA-1 binding (Figure 5.12).

5.2.2.4 Transcription factor binding around rs 66650371(TAC/---)

Previous ChIP-chip experiments demonstrated the presence of five GATA-1 binding peaks at block 2 of chromosome 6q; three of which were around the conserved GATA-1 binding motifs. Transcription factor prediction tools predicted another GATA-1 motif in close proximity to the conserved GATA-1 motif. I have investigated if any SNPs associated with elevated HbF levels altered these conserved or predicted motifs. There were no SNPs around the conserved GATA-1 binding site; however there was a three base pair deletion that was encompassed by the predicted GATA-1 binding motif. I

have designed oligonucleotides encompassing the predicted GATA-1 binding site (Figure 5.13) to confirm GATA-1 binding by EMSA experiments and to investigate if any transcription factor binding is altered by the deletion rs66650371. Initial experiments were performed under variable binding conditions to select the most appropriate binding condition for the oligonucleotides of interest.

Table 5.10 shows the alternative binding conditions that were tested. Table 5.11 shows the comparison of binding reactions. Initial results suggested that different transcription factor complexes are binding to the “wt” and “del” oligonucleotides, which were designed to encompass the three base pair deletion rs66650371 and predicted motifs for key transcription factors like GATA-1/GATA-2 TAL-1 and RUNX1. Results also demonstrated significant difference in the binding patterns of oligonucleotides with variable binding conditions (Figure 5.14).

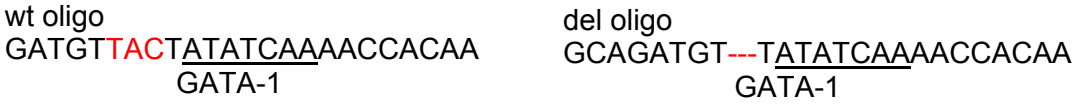


Figure5.13: EMSA oligonucleotides encompassing SNP rs66630371

Component	Binding condition 1	Binding condition 2	Binding condition 3
Ultrapure Water	9 µl	9 µl	8µl
10X Binding Buffer	2 µl	2 µl	2 µl
50% Glycerol	1 µl	1 µl	1 µl
MgCl2	1 µl	-	1 µl
Poly (dI•dC))	1 µl	1 µl	1 µl
1% NP-40	1 µl	-	1 µl
KCl	-	1 µl	-
EDTA	-	1 µl	1 µl
K562 nuclear extract	3 µl	3 µl	3 µl
Oligonucleotide	2 µl	2 µl	2 µl
Total Volume	20 µl	20 µl	20 µl

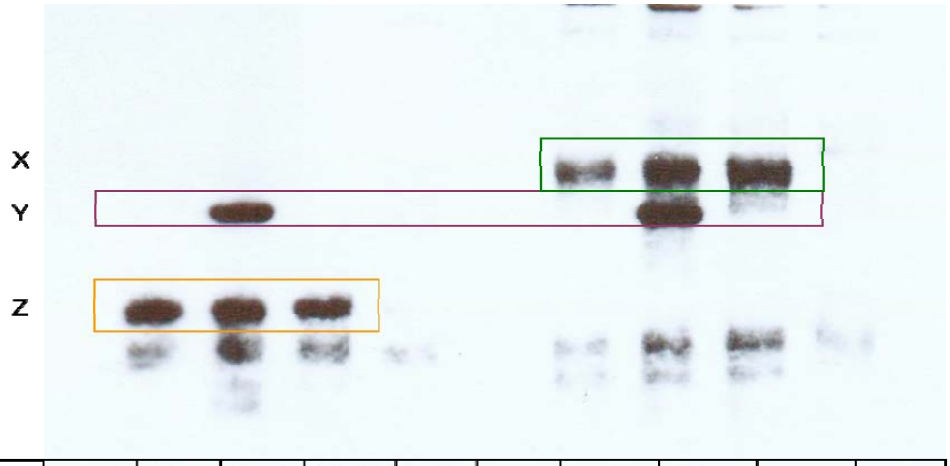
Table5-10: Three different binding conditions were prepared by alterations of KCL, EDTA, MgCl2, 1% NP-40 presence in the binding reaction.

	1	2	3	4	5	6	7	8	9	10
Extract	**-**	**K562**	**K562**	**K562**	**K562**	**-**	**K562**	**K562**	**K562**	**K562**
Oligo	**wt**	**wt**	**wt**	**wt**	**wt**	**del**	**del**	**del**	**del**	**del**
Competitor	**-**	**-**	**-**	**-**	**wt**	**-**	**-**	**-**	**-**	**wt**
Binding condition	**1**	**1**	**2**	**3**	**3**	**1**	**1**	**2**	**3**	**3**

Table 5-11: Binding Reactions for figure 5.14.
Three different binding reactions were prepared for each oligonucleotide .i.e. (wt and del) with variable binding conditions i.e. (Binding condition 1, 2, 3).

wt oligo
GATGTTACTATATCAAAACCACAA
GATA-1

del oligo
GCAGATGT---TATATCAAAACCACAA
GATA-1



Lane	1	2	3	4	5	6	7	8	9	10
Extract	-	K562	K562	K562	K562	-	K562	K562	K562	K562
Oligo	wt	wt	wt	wt	wt	del	del	del	Del	del
Competitor	-	-	-	-	wt	-	-	-	-	wt

Figure 5.14: Binding pattern of "wt" and "del" oligonucleotides with K562 nuclear extracts under variable binding conditions.

Protein-nucleic acid complex (X) is observed with "del" oligonucleotide (lane7, 8 and 9)) using all three binding conditions. Protein-nucleic acid complex (Z) is observed with the "wt" oligonucleotide (lane 2, 3 and 4) using all three binding conditions. Protein-nucleic acid complex (Y) is observed with both of the oligonucleotides (lanes3 and 8, respectively) under binding condition two only.

Lane 1 and Lane 6 are negative controls with no nuclear extracts. Cold competition with cold self-competitor competes out Z and X complexes with wt and del oligonucleotides, respectively (Lanes 5 and10, respectively, figure 5.14).The results suggest that binding condition two is the most appropriate condition for future experiments.

5.2.2.4.1 Cold Competition EMSA

Initial EMSA experiments demonstrated the presence of protein-nucleic acid complexes that are different and specific to the “wt” and “del” oligonucleotides. Based on the information obtained from *in-silico* predictions, experiments were performed using the commercial GATA-1 consensus sequences and the mutant GATA-1 sequences as cold competitors in order to investigate if any of the demonstrated protein-nucleic acid complexes is GATA-1 related. Table 5.12 shows the comparison of the prepared binding reactions.

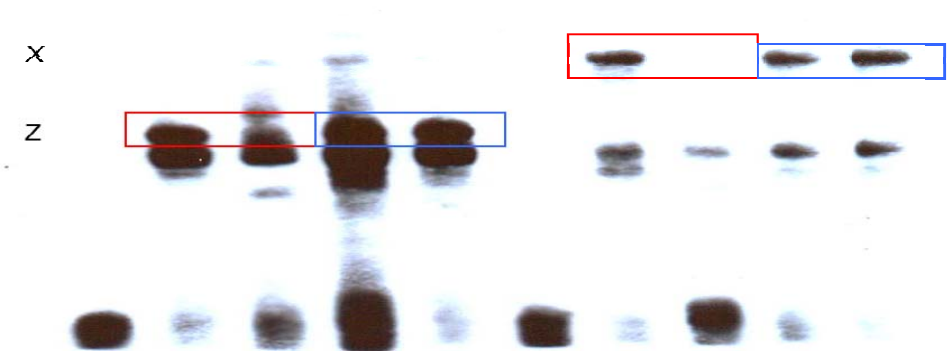
	1	2	3	4	5	6	7	8	9	10
Extract	**-**	**K562**	**K562**	**K562**	**K562**	**-**	**K562**	**K562**	**K562**	**K562**
Oligo	**wt**	**wt**	**wt**	**wt**	**wt**	**del**	**del**	**del**	**del**	**Del**
Competitor	**-**	**-**	**wt**	**GATA-1+**	**GATA-1-Mutant**	**-**	**-**	**del**	**GATA-1+**	**GATA-1-Mutant**
Condition	**2**	**2**	**2**	**2**	**2**	**2**	**2**	**2**	**2**	**2**

Table5-12:Binding Reactions for figure 5.15.Cold competition of “wt” and “del” oligonucleotides with GATA-1 + (consensus GATA-1 site), GATA-1 – (mutated consensus site) were performed.

All reactions were performed under binding condition 2; for competition 200 fold excess of unlabelled oligonucleotide was added to the reaction 5 minutes before addition of labelled oligonucleotides.

wt oligo
GATGTTACTATATCAAAACCACAA
GATA-1

del oligo
GCAGATGT---TATATCAAAACCACAA
GATA-1



Lane	1	2	3	4	5	6	7	8	9	10
Extract	-	K562	K562	K562	K562	-	K562	K562	K562	K562
Oligo	wt	wt	wt	wt	wt	del	del	del	del	Del
Competitor	-	-	wt	GATA-1+	GATA-1-Mutant	-	-	del	GATA-1+	GATA-1-Mutant

Figure 5.15: Cold competition patterns of "wt" and "del" oligonucleotides with GATA-1+, GATA-1- , "wt" and "del" unlabelled oligonucleotides (Table 5.11).
Specific protein-nucleic acid complex (X) is observed with "del" oligonucleotides (lane7) that is competed out by addition of excess self-competitor (lane 8). Specific protein-nucleic acid complex (Z) is observed with wt oligonucleotides (lane 2) that is competed out by addition of excess self-competitor (lane 3). Addition of GATA-1+ and GATA-1- (mutant) cold competitors with labelled "wt" and "del" oligonucleotides did not compete out complex X or complex Z , respectively.

In summary, although protein binding is demonstrated, the results failed to suggest that the specific protein-nucleic acid complex (X) and (Z) are GATA-1 related. Specific protein-nucleic acid complex (Y) observed in Figure 5.14 cannot be observed probably due to unstable nature of the complex. It is possible that this complex (Y) from figure 5.14 is the GATA-1 related complex. Significant difference in the binding patterns of the "wt" and "del" oligos is demonstrated and future experiments should focus on the characterization of "X" and "Z" protein-nucleic acid complexes.

5.3 Discussion

Two individuals, one homozygous (++) for the for the haplotype across the 6q QTL(blocks 1, 2 and 3) associated with high HbF and another homozygous for the absence of the SNPs associated with increased HbF levels, were selected for resequencing in order to investigate the SNPs that differ between a high F and a low F individual. *In-silico* prediction tools (matinspector and consite) were used to predict the transcription factor binding sites around these SNPs. Further comparison of predicted binding sites in the presence and absence of the SNPs and the alternative alleles was performed to investigate if the SNPs alter binding of the transcription factors. *In-silico* predictions highlighted some SNPs which altered the binding of key transcription factors like MYB, KLF1 and GATA-1 family.

The MYB transcription factor has a crucial role in the regulation of proliferation and cell lineage commitment of erythroid progenitor cells. Jiang et al showed that over expression of *MYB* in human erythroid progenitor cells results in down regulation of γ globin(Jiang et al 2006). These results were also supported by studies of Sankaran et al who showed that down regulation of *MYB* in human erythroid progenitor cells resulted in not only increased expression of γ-globin but also ε-globin(Sankaran et al 2011). The influence of MYB on HbF levels were further supported by Galarneau et al who showed that missense mutations in *MYB* strongly correlates with HbF. The study was performed in African Americans with sickle cell anemia (SCA) to characterize associations with fetal hemoglobin (HbF) levels using resequencing and genotyping. The group suggested that 3 missense variants in the *MYB* gene was associated with elevated HbF levels independently from variants in *HMIP* variants (Galarneau et al 2010).Another study also suggested that MYB transcription factor binds to 5' flanking region of the *MYB* gene, positively autoregulating the expression of the gene in order to

maintain high levels of MYB protein in cells that are proliferating (Nicolaides et al 1991). Levels of *MYB* clearly influences HbF hence, differential binding of the MYB transcription factor to candidate SNPs in the 6q intergenic region may indirectly control HbF through control of *MYB*.

Factors belonging to the KLF family are also crucial in the regulation of globin genes. KLF1 binds to the β-globin gene promoter sequence as well as the core of hypersensitive sites in the LCR which is essential for the β-globin gene expression. KLF1 activates the β –globin gene binding specifically to the CACCC box in the β-globin promoter. The KLF1 binding enhances the interaction between the β-globin promoter and the β-LCR, as a consequence of which, the interaction between the γ-globin promoter and the β-LCR decreases (Miller & 1993) facilitating the switch from γ-globin to β-globin gene expression in adults. Recent studies also suggested that KLF1 regulates the γ–globin to β–globin gene switch by activating transcriptional repressors of the γ –globin gene, such as BCL11A, TR2/TR4 and KLF10 (Borg et al 2010,Borg et al 2012).

Two different functional assays were performed to study differential binding of these factors at the predicted sites. Electrophoretic mobility shift assays (EMSA) were the first functional assays that were performed.
EMSA uses oligonucleotides encompassing the alternative alleles of the SNPs in binding reaction of nuclear extracts in which the putative transcription factor is expressed to demonstrate any differential transcription factor binding. Specificity of the binding can be investigated by cold competition assays as previously described.

Characterization of proteins that are present in the protein complexes requires additional steps. The electrophoretic “supershift” assay and assays that combine EMSA with western blotting or mass spectroscopy have been devised to allow

identification of nucleic acid-associated proteins. Supershift experiments proved to be challenging so we have used an alternative method to detect proteins present in the protein complex by competing the proteins out with unlabeled oligonucleotides that are known to be bound by proteins of interest.

Initial experiments were performed with oligonucleotides encompassing SNPs rs9389268 (A/G) and rs9376091 (T/C) that included the MYB (CAACCG) and KLF1 (GTGGG) binding motifs. Interpretation of EMSA results of rs9289268 proved very difficult because the two transcription factor binding sites that were only 4 base pairs apart and had to be included in the design of the oligonucleotides (Figure 5.5). It was hard to separate the binding sites because that would reduce the flanking sequence on one site leading to an artificial environment. On advice from Dr. David Garrick (WIMM, Oxford), we used the oligonucleotides encompassing both MYB and KLF1transcription factor binding rather than separating them.

Alternative combinations of binding reactions were tried out in order to detect the most appropriate one to improve sensitivity. Oligonucleotides suitable for 3 ^{I} end radioactive labelling were designed according to Dr. David Garrick's suggestions and Radioactive EMSA method was performed to get better results. Unfortunately Radioactive EMSA results were not successful.

The other problem with the EMSA experiments was the failure to super shift the protein-nucleic acid complex with antibodies to characterize the bands that were shown in the EMSA results. Different MYB and KLF antibodies were used but no consistent super shift was observed leading to the failure of characterization of the bands. This can result from the conformational changes of proteins after binding to oligonucleotides that prevent the binding of antibodies. Experiments were performed in which nuclear extracts are incubated with antibodies before the addition of labelled oligonucleotides in order to enable super shifts. However, addition of antibodies appears to disrupt the

protein-nucleic acid complexes that can be classified as a super shift yet the disruption was not specific to the antibodies of interest. Cold competition EMSA experiments were performed in order to characterize the demonstrated protein-nucleic acid complexes and results suggested that the protein-nucleic acid complexes around SNPs rs9389268 (A/G) and rs9376091 (T/C) are related to MYB and KLF1.

Binding patterns of old-g and old-a oligonucleotides with K562 nuclear extracts may suggest the formation of a trimeric complex which includes the labelled oligonucleotides together with the MYB, KLF1. Recent reports have described a transcriptional network within the erythroid cells whereby KLF1 and MYB activate each other. Bianchi et al suggested that MYB can regulate KLF1; and Stadhouders et al demonstrated that KLF1 binds to a site 84 kb from *MYB* gene transcription start site which coincides with a SNP strongly associated with high HbF (Bianchi et al 2010, Stadhouderset al 2012). Therefore, it is possible that MYB and KLF1 regulate expression of *MYB* gene. SNPs around the binding motifs may disrupt the complex or reduce its binding efficiency. This may explain the differential rate of proliferation and differentiation of erythroid progenitor cells from patients with the (++) haplotype. Other functional approaches should be taken into consideration in order to investigate the binding of these transcription factors in the chromosome 6q QTL.
EMSA experiments also focused on two other SNPs, rs9494142 and rs66650371, which disrupted the predicted GATA-1 binding motifs. Confirmation of differential binding of GATA-1 is crucial in understanding the regulatory activity of the region. Since GATA-1 can function both as a repressor or an activator of gene expression, differential binding of GATA-1 may lead to either upregulation or downregulation of flanking genes via its histone remodelling activity or through the complexes formed with other transcription factors. *In-silico* tools predicted GATA-1 binding which was supported by EMSA assays.However, although EMSA experiments confirmed GATA related complex formation with oligonucleotide encompassing predicted GATA-1 site

around SNP rs9494142; it failed to suggest any differential GATA-related binding. Cold competition experiments demonstrated that protein-nucleic acid complex is competed out with the addition of commercial GATA-1+ motif and the efficiency of competition is significantly less with GATA-1- (mutant). ChIP-chip experiments demonstrated GATA-1 binding around this region but initial focus was on the conserved GATA-1 site found in the region which is not altered by any SNPs. Potential alteration of GATA-1 predicted site can still be important since this site is in close proximity with conserved GATA-1 site and GATA-1 transcription factor which can function in 161immers as well. Disruption of one site may limit the binding of GATA-1 to the region *in vivo,* but this should be investigated by ChIp-qPCR experiments.

EMSA experiments performed with the oligonucleotide encompassing GATA-1 predicted site and the three base pair deletion (rs66650371) suggested differential binding patterns with the “wt” and “del” oligonucleotides. In order to characterize protein-nucleic acid complexes, cold competition assays were performed but the results failed to suggest that the differentially bound complexes are GATA-related. There are other key factor motifs that are present in the oligonucleotides such as: RUNX1, Hepatic nuclear factor 1 (HNF1) (Chapter 3) and future experiments should be performed in order to characterize these protein-nucleic acid complexes. Future experiments that combine EMSA with mass spectroscopy can be performed to identify proteins in the complex.

6 INVESTIGATION OF TRANSCRIPTION FACTOR BINDING AT THE 6Q23 LOCUS BY CHROMATIN IMMUNOPRECIPITATION (CHIP)

6.1 Introduction

Insilico predictions and EMSA experiments suggested presence of key transcription factors encompassing the SNPs associated with elevated HbF levels. EMSA experiments were performed in vitro with K562 nuclear extracts. I have performed ChIP experiments to investigate transcription factor binding using erythroid progenitor cells.

6.1.1 Chromosome 6q QTL

The QTL on the chromosome 6q23 was identified by linkage analysis in an extended Asian family with β-thalassaemia through the proband who had very mild thalassaemia intermedia despite being homozygous for β^0 thalassaemia with a complete absence of HbA ($\alpha_2\beta_2$). He was able to compensate for the lack of HbA due to the ability to produce HbF ($\alpha_2\gamma_2$); his steady state total haemoglobin ranged from 11-12 gm/ dl, all of which was HbF apart from small amount of HbA_2 ($\alpha_2\delta_2$). Family studies showed that several members of the extended family, with and without β thalassaemia had variable increases in HbF levels. Segregation studies showed that there was a genetic determinant for this innate ability to produce HbF and that the genetic determinants segregated independently of the β-globin gene complex. The extended family was large enough to perform linkage analysis, which mapped the HbF locus to a region spanning 1.5 Mb on chromosome 6q23. The 6q HbF QTL was validated in a large panel of European twins. Higher resolution genotyping refined the locus to a region of 79 kb between the *HBS1L* and *MYB* genes. The analysis revealed that this was a major QTL which accounted for 35% of total F-cell variance in the family (Thein et al 1994).

The 6q QTL exists as a set of single nucleotide polymorphisms (SNPs) spanning a region of ≈ 79kb that includes the intergenic region 5^{i} to *MYB and HBS1L and* 5^{i} part of the *HBS1L gene.*The SNPs are distributed in three linkage disequilibrium (LD) blocks referred to as *HBS1L –MYB* intergenic polymorphism *(HMIP)* blocks 1, 2, and 3. The 3 blocks completely account for the HbF variance attributed to the 6q QTL with SNPs in block 2 (24kb) showing the strongest effect and accounting for the majority of the variance.

Our group has shown that *MYB* and *HBS1L* are quantitative trait genes,with variable expression in the erythroid progenitor cells of healthy adults. Previous analysis of a number of erythroid cultures of individuals has revealed a clear correlation of elevated HbF with low levels of *MYB* and *HBS1L* expression (Jiang et al 2006).

Wahlberg et al *(*2009) identified Dnase I hypersensitive sites in *HMIP* block 2 suggesting the presence of regulatory elements in the *HBS1L-MYB* intergenic region further supported by the presence of erythroid specific transcripts. Further experiments were performed to analyse the region for transcription factor binding and chromatin modification by chromatin immunoprecipitation on microarray (ChIP-chip). The results indicated five GATA-1 binding sites in block 2 (figure 6.1). The use of microarrays allowed the investigation of GATA-1 binding to an extensive area of the 6q23 locus.

6.1.2 Chromatin immunoprecipitation (ChIP)

Chromatin immunoprecipitation (ChIP) is a powerful experimental approach that allows one to identify the proteins associated with specific regions of the genome. With the appropriate antibodies, it can be used to locate both non-histone proteins and histones carrying specific covalent modifications, such as acetylation, phosphorylation, or methylation. Recent developments in techniques such as quantitative real-time PCR and DNA microarrays have enabled efficient analysis of ChIP material.

There are two general methodologies for carrying out ChIP experiments. One (XChIP) uses chromatin fixed with formaldehyde and fragmented by sonication. The other (NChIP) uses native chromatin prepared by nuclease digestion of cell nuclei. Both methods have advantages and disadvantages, the method used will depend, among other things, on the aims of those experiments and the starting material.

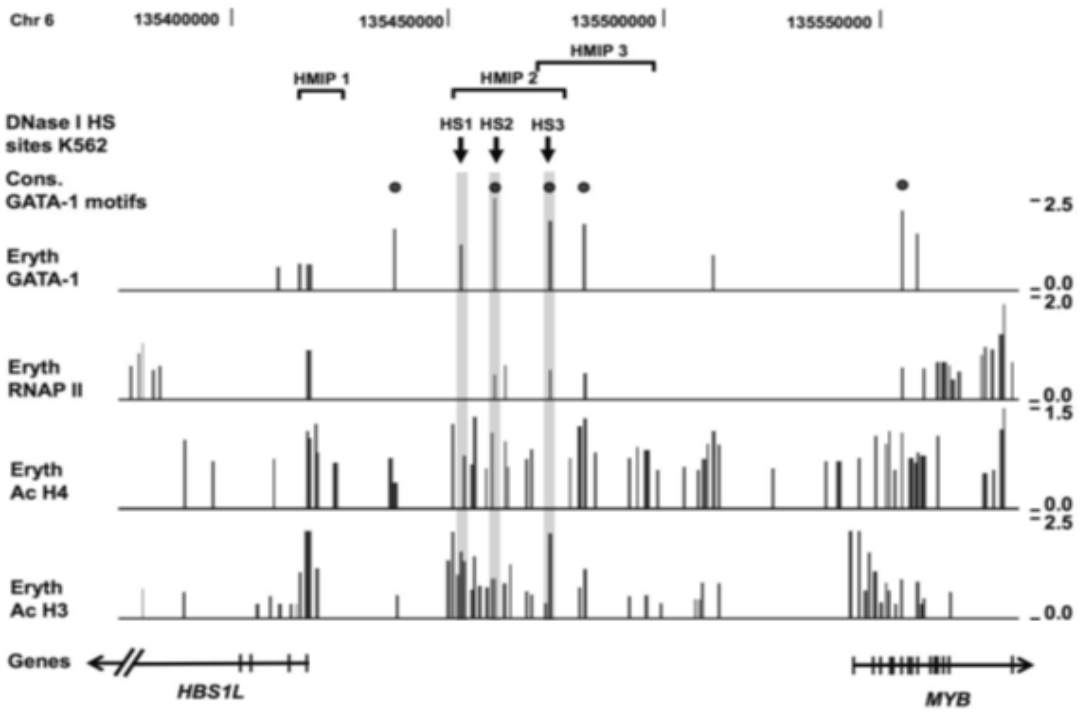


Figure 6.1: ChIP-chip data for the HBS1L-MYB intergenic region.
Results from ChIP-chip experiments for GATA-1, RNA polymerase II, AcH4, and AcH3 in primary human erythroid precursors are presented. Covering 210-kb of 6q locus (position chr6: 135 373 000-135 582 000) including part of the *HBS1L* gene, the intergenic region, and the *MYB* gene. The *HMIP* block 1, 2, and 3 are shown above indicated by the horizontal brackets. Also shown are the conserved GATA-1 motifs and DNase I hypersensitive sites (as identified in hemin induced K562 cells by Wahlberg et al2009).Three DNase hypersensitive sites and five conserved GATA-1 motifs are highlighted on the figure which coincides with GATA-1 binding. Eyth Ac H4 and H3 show the histone acetylation patterns and RNAP II shows RNA polymerase activity in the region.

I have used the XChIP approach in this project. It is possible to start with smaller number of cells using XChIP, which enables us to perform additional ChIP experiments with different antibodies whose target proteins are expressed at similar stages during erythropoiesis. The main problem with this method is the small yield of precipitated chromatin limiting the number of real-time PCR experiments that can be performed and necessitating a whole genome amplification (WGA) step before analysing material by microarrays.Material used for ChIP experiments was harvested from human erythroid progenitor cells, derived from primary cultures at a time when our protein of interest is expressed. To obtain an idea of the appropriate time of harvest, we generated expression profiles for each target protein to be studied.

Primary human erythroid cells were cultured from whole blood using the method adapted from Van den Akker et al as detailed in Chapter 3. RNA was extracted from cells at different days of culture and reverse transcribed into cDNA. Using primers specific for each target gene, expression was measured by RT-PCR and we used this information in order to estimate the optimal day to harvest cells for ChIP experiments. On harvesting, cells were cross-linked using 1% formaldehyde; this step fixes the proteins to the target DNA sequence in their native state. The material is then sonicated to shear the fixed fragments between 200-1000bp (NimbleGen array's user guide for ChIP applications). The material was then immunoprecipitated with an antibody to our specific protein of interest with a mock antibody (rabbit IgG) in parallel, as a negative control. After an overnight incubation, any excess antibody is washed away and the antibody-protein-DNA material is reverse cross-linked, this time detaching the protein from its target DNA sequence. At the end of this step, we are left with DNA that has been specifically targeted by our protein using a specific antibody. The DNA is precipitated and quantitative PCR experiments are performed to quantify transcription factor binding to areas of interest (Figure 6.2). The enrichments are calculated relative to the IgG negative control using delta CT method.

We have performed experiments following two different approaches to measure differential binding of transcription factors. Initially, we have comparedtranscription factor binding to targetsin HPFH +/+ vs. HPFH -/- erythroid chromatin. Next, we have performed allele specific ChIP (HaploChIP) experiments with HPFH +/- individuals, heterozygous for the SNPs of interest, to confirm the differential transcription factor intraindividually.

Each approach was performed with different analysis methods of immunoprecipitated material. To compare the binding with HPFH +/+ and HPFH -/- erythroid chromatin we have performed ChIP-qPCR experiments, where the DNA is precipitated and quantitative PCR experiments are performed in order to quantify transcription factor binding to areas of interest (Figure 6.2). The enrichments are calculated relative to the IgG negative control using delta CT method. For further comparisons, enrichments for each target are normalized to relevant positive control to enable us compare the transcription factor binding between individuals with different genotypes. The detailed ChIP-qPCR protocol can be found in Chapter 2 (Materials and Methods).

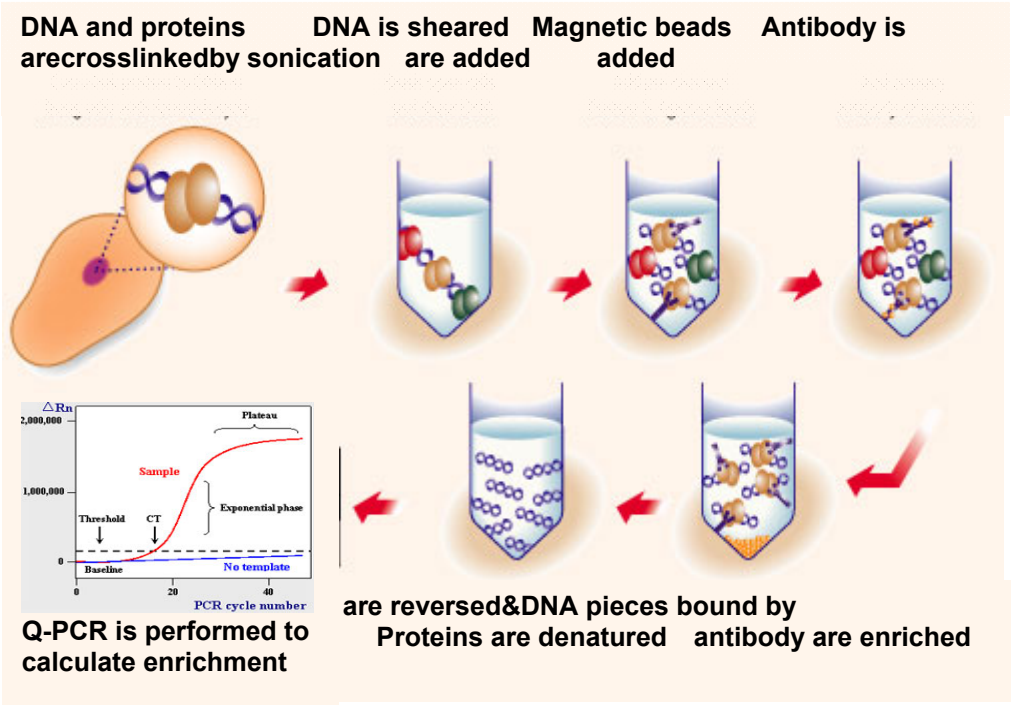


Figure6.2: Summary of the steps in ChIP-qPCR (The figure was adopted from vinci-biochem).

The second approach, allele specific ChIP or Haplochip enabled the analysis of differential binding of transcription factors to the two alleles representing a SNP in HPFH +/- erythroid chromatin. This approach enables us to overcome the limitations of ChIP-qPCR experiments performed with different individuals of different genotypes. Haplochip uses cells from the same individual who is heterozygous for a SNP of interest i.e. HPFH +/- individuals.

Different approaches of analysis depending on the nature and location of the SNPs can be applied to quantitate the biased binding of transcription factors to SNPs. These include: restriction digestion, snapshot and sequence analysis.

6.1.3 Transcription Factors at Chromosome 6q QTL

Transcription factors play a crucial role in gene expression. They bind to specific recognition sequences or binding motifs and help in the activation or repression of genes acting as *trans* regulators. Several transcription factor-binding motifs have been characterized within the β-globin locus and are essential for the formation of the active locus and activation and control of individual genes. The core regions of the hypersensitive sites in the LCR are particularly rich in such motifs and include binding sites for the erythroid specific transcription factors KLF1, GATA-1 and NF-E2 as well as ubiquitous factors such as SP1 (Goodwin et al 2001; Ho & Thein 2000).

The chromosome 6q QTL also has several transcription factor-binding motifs of GATA-1 and MYB transcription factors. Factors belonging to KLF family: SP1, BKLF, KLF1 are also present in the 24 kb region of block 2 where association with increased fetal haemoglobin is strongest (Thein et al 2007). Some of these transcription factors were discussed in Chapter 1 in detail because of their crucial roles. Previous work by our group demonstrated the presence of GATA-1 binding sites in the 6q intergenic

region using ChIP-Chip assay (Wahlberg et al2009). *In-silico* experiments performed also revealed MYB, KLF1 and GATA-1 binding sites in close proximity with SNPs that are associated with HbF.

GATA-1 is an erythroid specific transcription factor, required for globin gene regulation and erythroid cell maturation. GATA-1 can function both as a repressor or an activator of gene expression and the function is partly dependent on its interactions with other proteins. GATA-1 can interact with itself as well as other transcription factors including KLF1 and SP1 and can form large multiprotein complexes with TAL1, E47, LDB1, and LMO2 (Crossley et al1995 ; Merika &Orkin et al 1995; Wadman et al 1997; Morceau et al 2004; Cheng et al 2009).

The GATA-1 recognition sequence (T/A) GATA (A/G) is found in the globin-gene promoters as well as in the core sequences of the hypersensitive sites of the LCR. It is also involved in the establishment of the active chromatin domain in the locus and the formation of hypersensitive sites (Orkin et al 1992; Pomeratz et al 1998; Vakoc et al 2005; Layon et al 2007). GATA-1's potential as a gene regulator and important regulatory roles in erythropoiesis, makes it an ideal candidate to study transcriptional activity at the 6q23 locus.

Another transcription factor of interest was MYB. The *MYB* gene encodes a protein of 72k Da comprising 640 amino acids with an isoform of 89 kDa (cMYB-9A) caused by a splice variant which leads to an insertion of 363 base pairs (bp) between exon 9 and 10 (exon 9A). The protein is expressed in most haematopoietic tissues and is highly expressed in immature haematopoietic cells (Gonda et al 1982, Westin et al 1982).

MYB transcription factor plays crucial roles in regulation of proliferation and cell lineage commitment of erythroid progenitor cells. It is required for the maintenance of cell proliferation by exerting a direct role in cell cycle control. *MYB* is highly expressed in

immature haematopoietic cells and is down regulated as the cells differentiate. Down-regulation of *MYB* is required for terminal differentiation (Westin et al 1982, Oh et al 1999).

Krüppel-like factor 1, KLF1, is an erythroid cell-specific transcription factor, homologous to the Krüppel family of transcription factors which have roles in cell proliferation, differentiation and survival. KLF1 contains three zinc fingers at the C-terminus which bind to a CACCC sequence. CACCC sequences are repeated in erythroid enhancers and promoters, including the β-globin gene promoter. This sequence is noted as a site of point mutations that give rise to β-thalassaemia.*Klf1* knockout mice die from anemia supporting the role of KLF1 during erythropoiesis (Miller & Bieker, 1993, Bieker et al 2005). Recent studies suggested that KLF1 has a direct role in globin regulation through the activation of *BCL11A* (Borg et al 2010).

6.2 Results

6.2.1 DNA sonication

Following crosslinking of primary human erythroid cells and cell lysis, the material was sonicated to shear the DNA. Electrophoresis of 1µl of the sonicated chromatin was performed on 1.5% acrylamide gel to confirm the size of the DNA fragments (Figure 6.3, lanes 2-3). Figure 6.3 shows that chromatin was successfully sheared to 100-600bp.

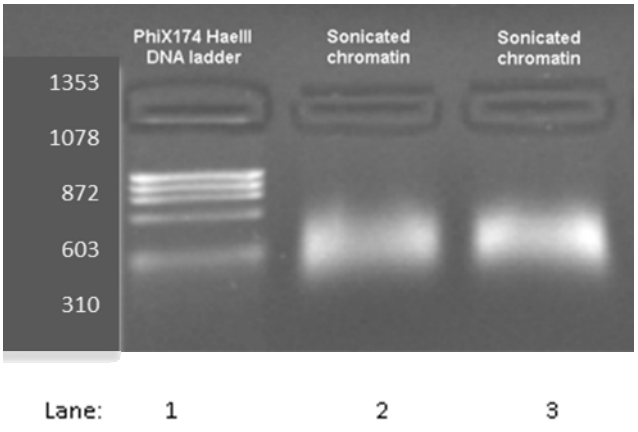


Figure6.3: Electrophoresis of two sets chromatin sonicated for 15 rounds at power 2 for 10 seconds (lanes 2 and 3) with PhiX174 HaeIII DNA ladder (lane 1) on 1.5 % w/v agarose gel.

6.2.2 GATA-1 Binding at chromosome 6q QTL

Experiments were performed to analyse the region for transcription factor binding and chromatin modification by chromatin immunoprecipitation on microarray (ChIP-chip). The results indicated five GATA-1 binding sites in block 2 (figure 6.5). The use of microarrays allowed the investigation of GATA-1 binding to an extensive area of the

6q23 locus. ChIP-chip is not a quantitative method so ChIP-qPCR experiments were performed to quantify GATA-1 binding to the previously revealed peaks.

6.2.2.1 Control test for GATA-1 enrichment

Following confirmation of the chromatin size, we proceeded to immunoprecipitation using an anti-GATA1 antibody (M-20 sc1234, Santa Cruz). After reverse cross-linking and purification of the ChIP DNA, a small amount was tested for enrichment by SYBR green RT-PCR. Hypersensitive site 40 (HS-40) upstream of the α-globin locus on chromosome 16 (*α-HS40)* and hypersensitive site 2 (HS-2) at β-globin LCR on chromosome 11 are known GATA-1 binding sites. Both sites were used as positive controls in GATA-1 ChIP experiments. This step is a check point to confirm if ChIP with the target antibody has been successful.

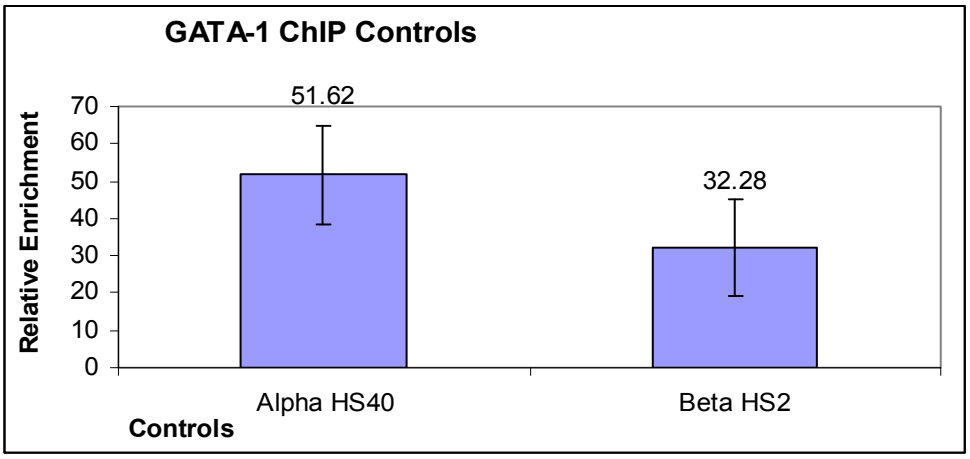


Figure6.4: SYBR-Green RT-PCR for *Alpha-HS40* and *Beta-HS2* enrichment was performed to confirm GATA-1 ChIP (using antibody GATA-1 M-20 sc-1234, Santa Cruz).

Shown as fold enrichment, normalized to NFM as a housekeeper gene and IgG (ChIP with mock antibody). Enrichment results are Mean values of 5 biological replicates +/- standard deviation.

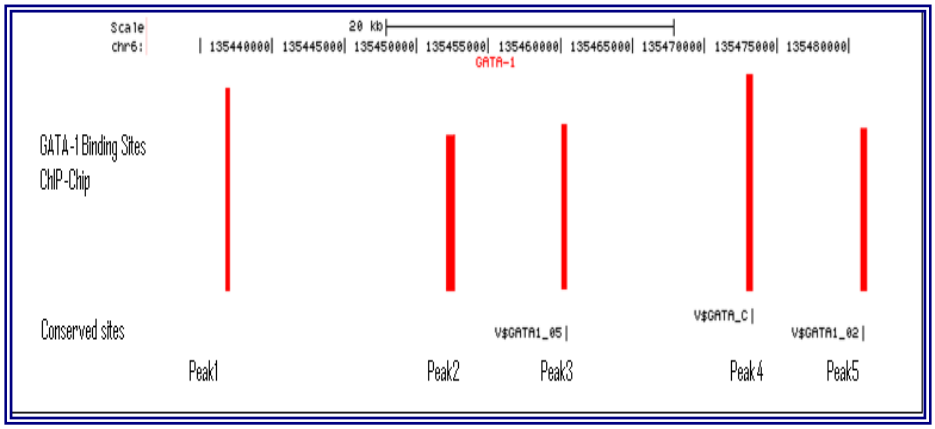
6.2.2.2 Quantitative GATA-1 binding at the 6q23 locus (HMIP)

Previous microarray based experiments (ChIP-chip) in the region demonstrated the presence of GATA-1 binding at specific positions at the 6q23 locus (*HMIP*). Although the use of microarrays would allow us to see a more extensive area of the 6q23 locus, these results are not quantitative and should be validated by qPCR. Hence, we designed primers for ChIP-qPCR assays to quantify the detected sites at the 6q23 locus (*HMIP*) block 2 (Figure 6.5 A).

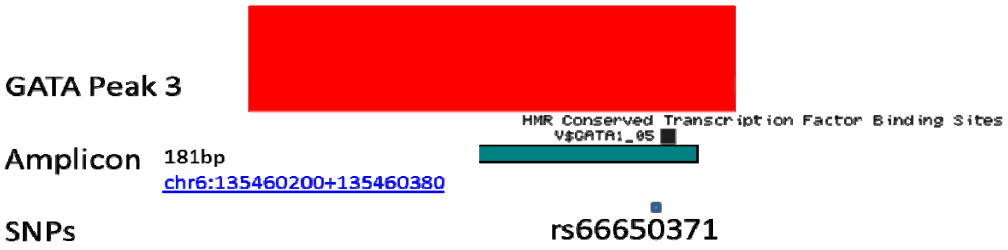
Tracks of the GATA-1 peaks at the 6q23 locus and surrounding area prepared by fellow students Karin Wahlberg and Kiran Jawaid were uploaded and viewed on the online UCSC genome browser (genome.ucsc.edu). This provides a more detailed view of the region, with the flexibility of viewing SNP tracks, conserved GATA-1 sites, hypersensitive regions, other transcription factor sites, as well as many other tracks available, all in the same view. The tracks of high scoring SNPs prepared by Stephen Menzel was uploaded together with GATA-1 tracks to enable us design ChIP-qPCR primers at the right area of GATA-1 peaks which covers an area of ≈ 500bp. Primers were designed for the sites that were consistently detected on array-based experiments and three of them coincided with conserved GATA motifs. Experiments were performed to optimise conditions for quantitative PCR on GATA-1 ChIP material from healthy individuals. Mock IgG method was used in calculation of the enrichment on target sites. This method depends on normalization with a mock IgG: Fold enrichment = $2^{(ct\ IgG - ct\ ab)}$.

The analysis suggested strong GATA-1 binding to all peaks with relatively higher enrichments observed at peak three and peak four. Two different primer sets were used for peak three and there is a significant difference in terms of enrichment between the two primer sets. The set amplifying Peak 3-2(Primer 3-2) was selected for future experiments (Figure 6.5 B).

A



B



C

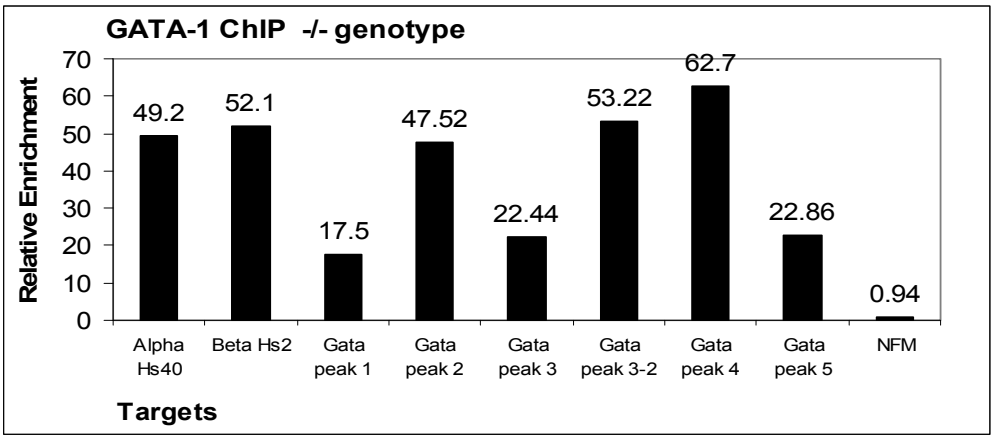


Figure6.5. Quantification of ChIP-Chip experiments

A: Gata-1 Binding Sites from previous ChIP-chip experiment and conserved GATA motifs coinciding with the binding sites. B: General designs of qPCR primers.ChIP-qPCR amplicons encompass the conserved GATA-1 motif and key SNP in the region. C: SYBR-Green RT-PCR was performed. Enrichments relative to IgG (rabbit immunoglobin G) were calculated to confirm quantitative GATA-1 binding to targets.

Alpha Hs40 and Beta Hs2 were used as positive control and *NFM* was used as a negative control.

6.2.2.3 Investigation of differential GATA-1 binding at the 6q23 locus using ChIP-qPCR in HPFH +/+ vs HPFH -/- erythroid chromain

Previous microarray based experiments (ChIP-chip) region suggested differential GATA-1 binding at specific positions (Peak 4) at the 6q23 locus (*HMIP*). These experiments were performed with material from individuals that were designated as HPFH +/+ and HPFH -/- . The HPFH +/+ individual is homozygous for the SNP alleles associated with high HbF across the 6q QTL (blocks 1, 2 and 3) while the HPFH -/- individual is homozygous for the absence of the SNP alleles associated with increased HbF levels .ChIP-chip experiments suggested differential binding of GATA-1 to one region. To further investigate this region, quantitative ChIP-qPCR was used to validate the differences suggested by the ChIP-chip experiments. Quantitative ChIP-qPCR was also used to identify any other potential differences in GATA-1 binding between the HPFH -/- and HPFH +/+ individuals.

Primary erythroid cells were cultured from whole blood taken from three paired sets of HPFH +/+ and HPFH -/- individuals. The HPFH +/+ individuals were members of the same family in which the QTL was first identified. One +/+ individual donated blood twice. ChIP-qPCR was carried out as previously described using the same anti-GATA-1 antibody (M-20 sc1234, Santa Cruz). To ensure that we harvest the erythroid progenitor cells from the +/+ and -/- individuals at equivalent stages of GATA-1 expression, we initially obtained a profile of GATA-1 expression in cells from the individuals of both genotypes. The expression profiles showed that GATA-1 is highly expressed at day 7 with no significant difference between individuals of different genotypes, hence day 7 would be optimal for harvesting of the erythroid progenitors.

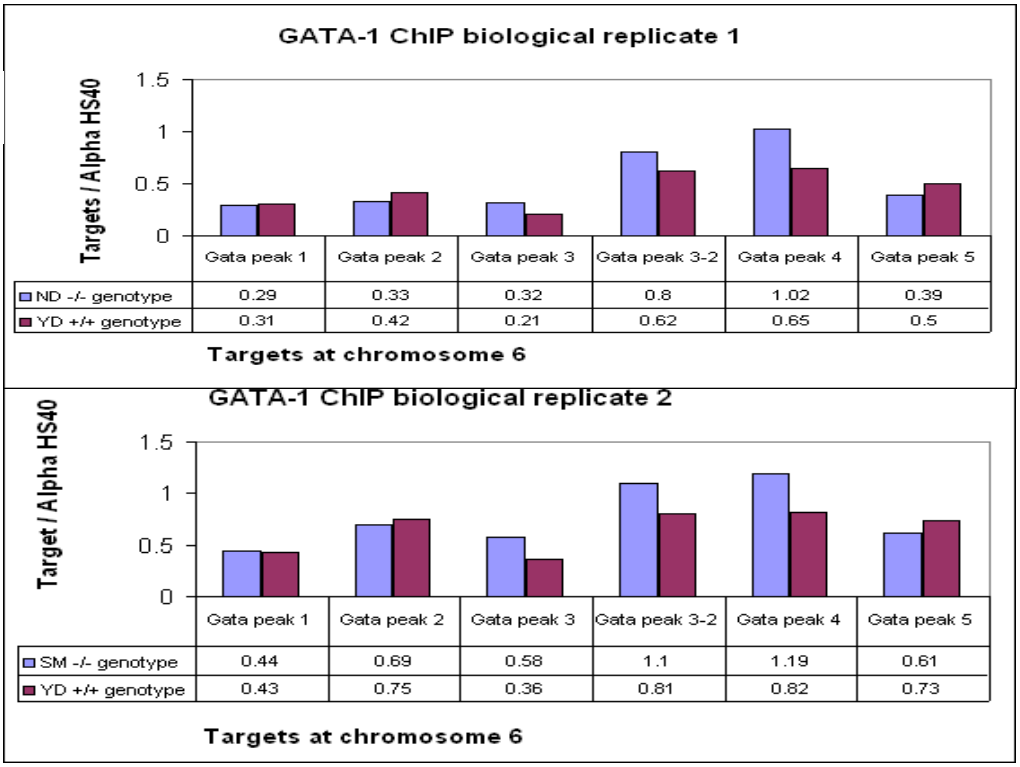
As validation, the cells were also closely monitored by FACS analysis as described in chapter 3. In performing ChIP on erythroid progenitor cells at similar stages of expression, we ensure any difference observed in GATA-1 binding is due to the alteration of binding by the SNPs and not because of differential expression of GATA-1.

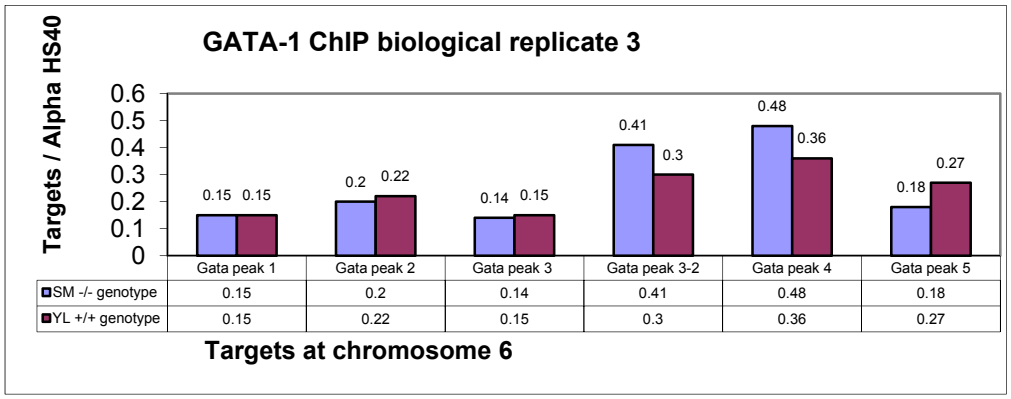
SYBR-Green RT-PCR was performed with material from six erythroid cultures (three HPFH -/- individuals and two HPFH +/+ individuals). Two technical experiments for each culture were performed. Enrichments relative to IgG (rabbit immunoglobin G) were calculated to confirm quantitative GATA-1 binding to targets. Results are normalized to positive control Alpha HS40 enabling us to cross compare GATA-1 binding to the same target in different individuals.Results confirmed the differential binding to GATA-1 peak four as suggested by the ChIP-chip experiments and also revealed differential GATA-1 binding to peak three. The results are consistent with the same binding pattern observed in all three biological replicates (Figure 6.6). Peaks 3 and 4 coincide with conserved GATA-1 sites as well as DNase hypersensitive sites (HS2 and HS3) which increase the likelihood of causative variants being in these regions.

6.2.2.4 Investigation of allele specific binding of GATA-1 to high scoring SNPS at the 6q23 locus in HPFH +/- erythroid chromatin

ChIP-qPCR experiments performed with (+/+) and (-/-) chromatin suggested differential GATA-1 binding at GATA-1 peak 3 and peak 4. Resequencing of the HPFH +/+ and HPFH -/- individuals revealed SNPs that were in close proximity to GATA-1 binding motifs (figure 6.5). We wanted to target individuals that are heterozygous for the key SNPs to investigate differential binding of GATA-1 to the minor allele (+) and major allele (-) intra-individually by allele specific ChIP. Healthy unrelated individuals were recruited and genotyped for the SNPs of interest (rs9494142 and rs66650371).

A





B

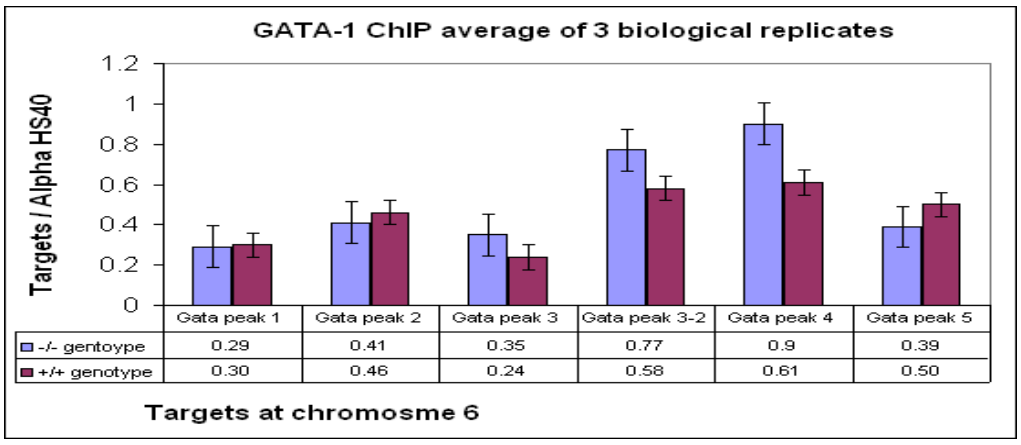


Figure6.6: Differential GATA-1 binding at the 6q23 locus. A: SYBR-Green RT-PCR was performed with material from six individuals (three -/- and three +/+/).
Enrichments relative to IgG (rabbit immunoglobin G) were calculated to confirm quantitative GATA-1 binding to targets. Results are normalized to positive control Alpha HS40 to enable as cross compare GATA-1 binding to same targets with different individuals. B: Mean of GATA-1 binding of replicates is presented with error bars for standard deviation.T-test was performed (n: 3) and suggested that there is no significant difference between two groups.

Cell cultures were performed with erythroid progenitor cells isolated from peripheral blood of individuals who were heterozygotes for the candidate SNPs.

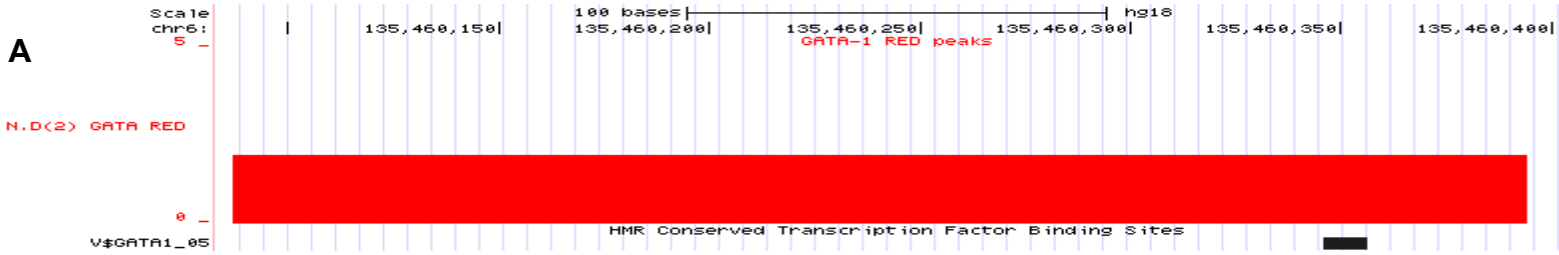
Erythroid progenitor cells were harvested and cross-linked at the stage of peak GATA-1 expression. Cross-linked material was immunoprecipitated with the same GATA-1 (M-20) antibody and chromatin was eluted as previously described. Eluted chromatin was tested for enrichment by SYBR green RT-PCR. Enrichment of positive control Alpha HS40 andGATA-1 sites of interest (peak 3 and peak 4) was calculated (Figure 6.7).

Different methods of analyzing the differential binding were applied to each SNP because of their different nature. SNP rs66650371 was a deletion which disrupts the cutting site of restriction enzyme *MaeIII.* Material immunoprecipitated with GATA-1 and input DNA were used as templates and a 91bp region encompassing the deletion was amplified by standard PCR. Restriction enzyme digestion was performed and agarose gel electrophoresis of digests was performed on 3% w/v agarose gel. Ratio of pixel density of undigested to digested material from input was used as a normalizer and differential binding of GATA-1 to each allele was quantified by calculating the pixel density of bands by fellow colleague Ralph Stadhouders (Figure 6.8).

SNPs rs9494142 (T/C) on the other hand, does not involve any restriction enzyme cleavage site. The differential binding to each allele was analyzed by snapshot method as previously described in Materials and Methods. The region encompassing polymorphism was amplified by standard PCR. Snapshot was performed with ChIP material and input as described before. Four biological replicates were used in the experiments and the T to C ratio of ChIP material was calculated and normalized to input (Figure 6.9).

Details of each protocol can be found in materials and methods.

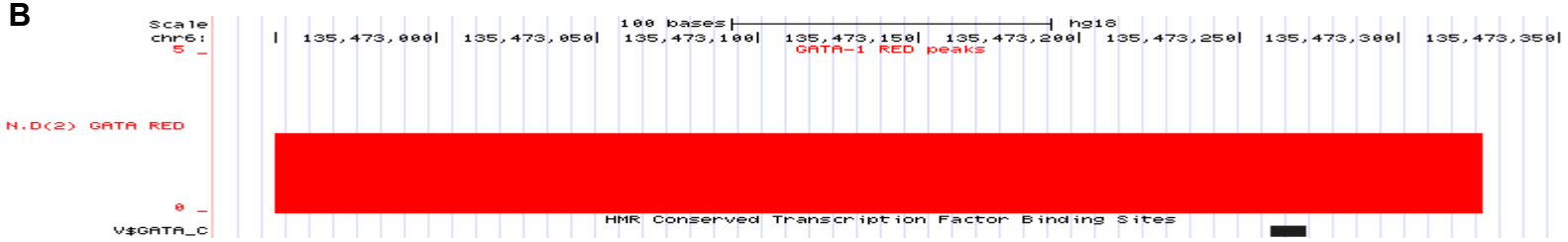
A



rs66650371 (TAC/---) GATA-1 Conserved GATA-1

CTCTACTGACACAACATCAGGATTAAATTCACTCTGGACAGCAGATGTTACTATATCAAAACCACAAAATGTTATCAGGGCGGTTCCCTCAGAAGACACT

B



Conserved GATA-1 rs9494142 (T/C) GATA-1 KLF1

TTCTGCAACTTCTAGATAAGTAAATTTTTTTGATTTATCAGGAAGTGTCTTTGGTCTCTCAGTCAATTCGATTCTACTACTGACATTTATCAACATGGTGGGTGTGAT

Figure 6.7: A: Gata-1 Peak 3 from previous ChIP-chip experiments. Peak 3 is 84 kb away from the MYB transcription start site (-84kb). Conserved GATA motifs, predicted transcription factor motif s and high scoring SNPs are found in close proximity. B: Gata-1 Peak 4 from previous ChIP-chip experiments. Peak 4 is 71 kb away from the MYB transcription start site (-71kb) .Conserved GATA motifs, predicted transcription factor motif s and high scoring SNPs are found in close proximity.

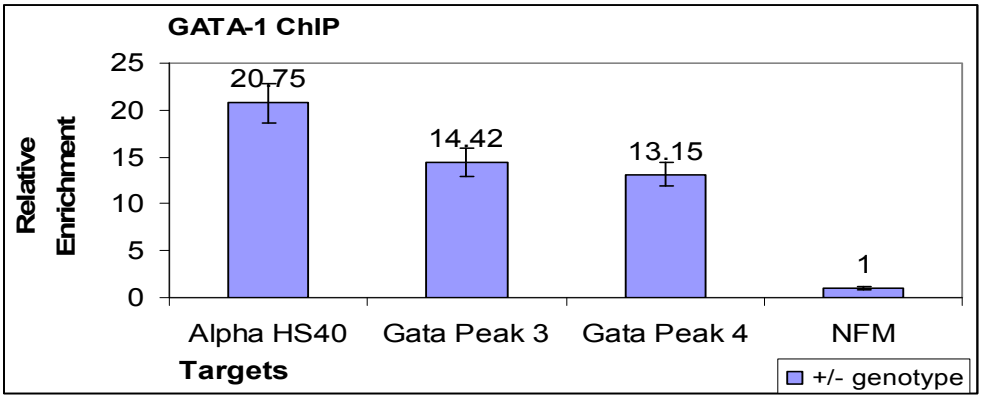
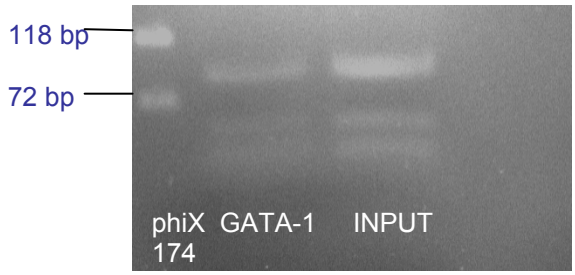


Figure6.8: SYBR-Green RT-PCR was performed to confirm GATA-1 enrichment of positive control Alpha HS40 and target sites Gata-1 peak 3 and 4 .
Shown as fold enrichment, normalized to *NFM* as a housekeeper gene and to IgG (ChIP with mock antibody). Enrichment results are Mean values of 4 biological replicates +/- standard error.



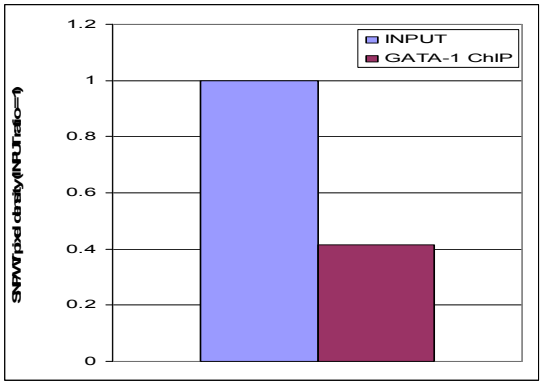
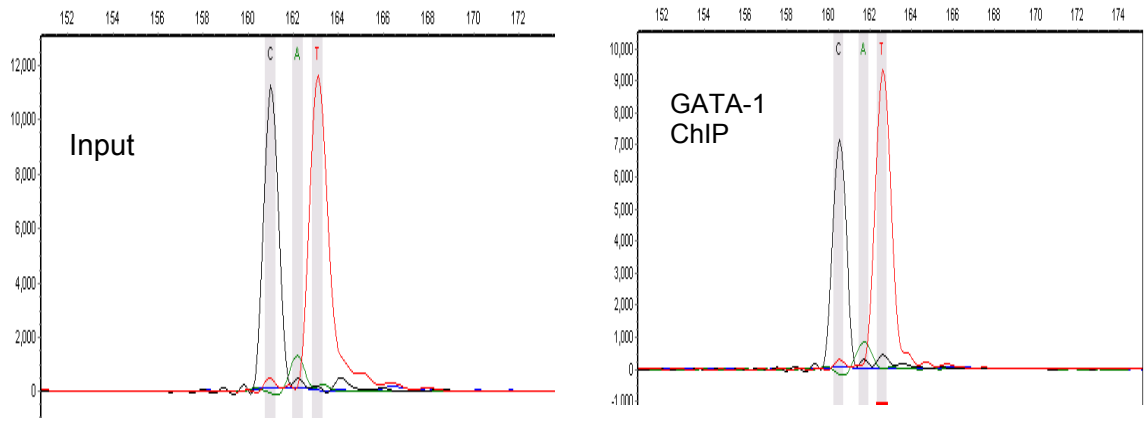


Figure 6.9: Allele specific GATA-1 binding to rs66650371.
A: Allele specific binding of GATA-1 to major allele (TAC) and minor allele (deletion) was quantified by Mae III digestion. ChIP with GATA-1(m-20) and input were digested and agarose gel electrophoresis was performed. B: Pixel density of bands with ChIP material was calculated and normalized to input.

A)



B)

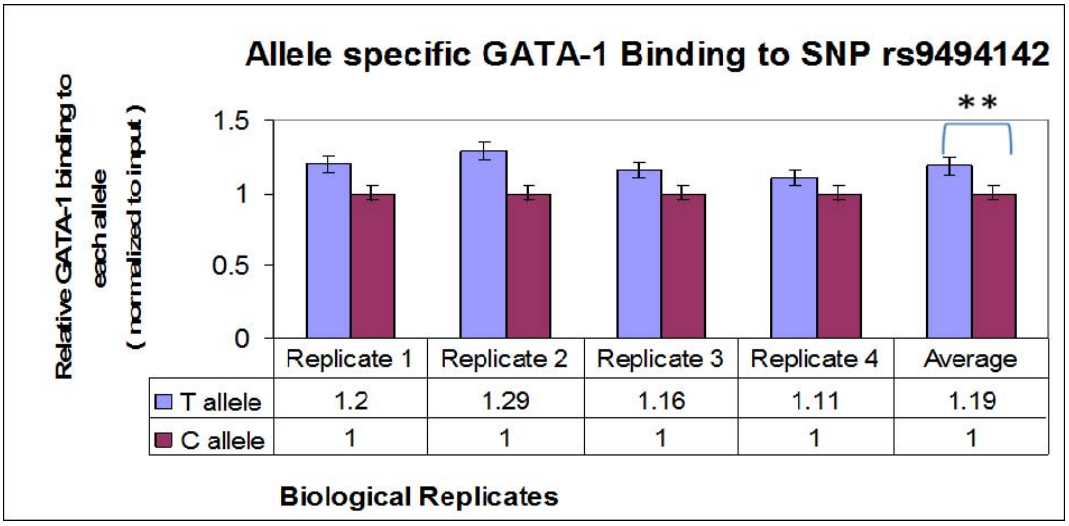


Figure.6.10: Allele specific GATA-1 binding to rs9494142.
A: Snapshot was performed with chromatin immunoprecipitated with GATA-1(m-20) and input. Ratio of GATA-1 binding to major allele T and minor allele C was calculated by using gene marker program which provides the peak height of both alleles. Calculation were done with both chip material and input for each individual.

B: Experiments were performed with four biological samples that were heterozygotes for the SNPs of interest .T to C ratio of chip material was calculated and normalized to input. T-test was performed (n: 4) and suggested that there is significant difference between twoalleles (p:0.002).

The results suggested weaker binding of GATA-1 to the minor alleles that are associated with elevated HbF. These results are consistent with the results observed earlier with the experiments performed with HPFH +/+ and -/- individuals. The main difference and the main advantage of this approach over ChIP-qPCR experiments is that the differential binding is internally controlled, delineating the key variants involved.

6.2.3 MYB binding at chromosome 6q QTL

ChIP experiments were performed to confirm *in-silico* experiments that suggested presence of MYB binding sites at chromosome 6q.

6.2.3.1 Control test for MYB enrichment

Following confirmation of the chromatin size, we proceeded to immunoprecipitation using an anti-MYB antibody (clone 1:1, Millipore). After reverse cross-linking and purification of the ChIP DNA, a small amount was tested for enrichment by SYBR green RT-PCR. Cathepsin G promoter (ctsg) is a known MYB binding site and was used as positive control in MYB ChIP experiments. This step is a check point to confirm if ChIP with the target antibody has been successful.

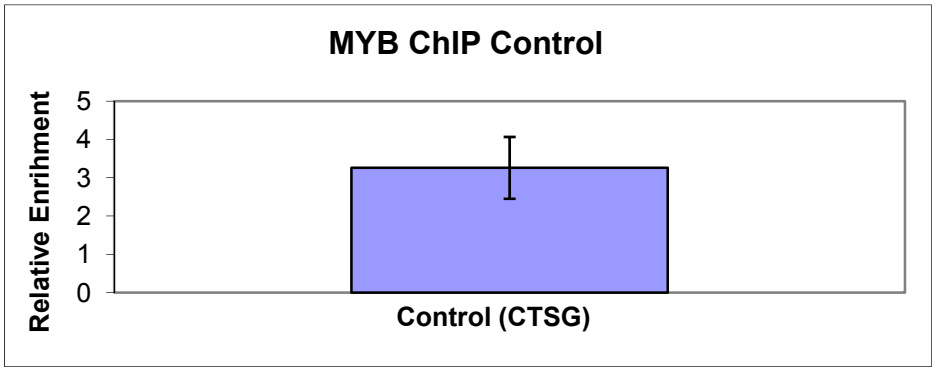


Figure6.11: SYBR-Green RT-PCR for CTSG enrichment was performed to confirm MYB ChIP (using antibody MYB clone 1:1, Millipore).

Shown as fold enrichment, normalized to *NFM* as a housekeeper gene and IgG (ChIP with mock antibody). Enrichment results are Mean values of 5 biological replicates +/- standard deviation.

6.2.3.2 Quantitative MYB binding at the 6q23 locus (HMIP)

Previous *in-silico* experiments demonstrated the presence of MYB binding motifs at specific positions at the 6q23 locus (*HMIP*) (See Chapter 3). The presence of some of the MYB binding sites was validated by EMSA experiments (Chapter 4). Although EMSA experiments are informative the results were obtained *in vitro* from K562 cell lines. We wanted to validate MYB binding to predicted sites in the 6q23 locus in erythroid progenitor cells by ChIP-qPCR. Primers were designed for the sites that were strong candidates based on *in-silico* experiments. Experiments were performed to optimise conditions for quantitative PCR on MYB ChIP material from healthy individuals. Mock IgG method was used in calculation of the enrichment on target sites: This method depends on normalization with a mock IgG: Fold enrichment = $2^{(ct\ IgG-ct\ ab)}$.

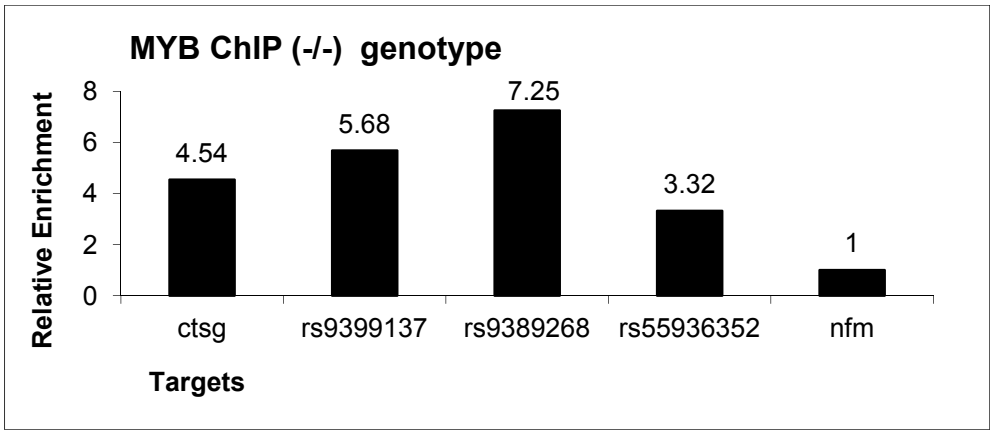


Figure6.12: MYB Binding at chromosome 6q: SYBR-Green RT-PCR was performed.
Enrichments relative to IgG (rabbit immunoglobin G) were calculated to confirm quantitative MYB binding to targets. CTSG was used as positive control and *NFM* was used as a negative control.

Strong MYB binding was observed around SNPs rs9399137 and rs9389268 although enrichments observed are significantly less compared to GATA-1 ChIP, and this is probably related to the antibody (Figure 6.12).
A number of different MYB antibodies were tried in the experiments, the MYB antibody 1:1 Millipore gave the best enrichments for chromatin immunoprecipitation experiments.

6.2.3.3 Investigation of differential MYB binding at the 6q23 locus by ChIP-qPCR in HPFH +/+ vs HPFH -/- chromatin

Previous *in-silico* experiments suggested that MYB binding motifs are altered by certain high scoring SNPs at specific positions at the 6q23 locus (*HMIP*). To further investigate this region with individuals that are HPFH +/+ or HPFH -/- quantitative, ChIP-qPCR was used to validate the differences suggested by the *in-silico* experiments.

Erythroid progenitor cells were obtained from primary cultures of three paired sets of HPFH +/+ and -/- individuals. The HPFH +/+ individuals were members of the same family in which the QTL was first identified. One +/+ individual donated blood twice. ChIP-qPCR was carried out as previously described using the same anti-MYB antibody (1:1 clone, Millipore).

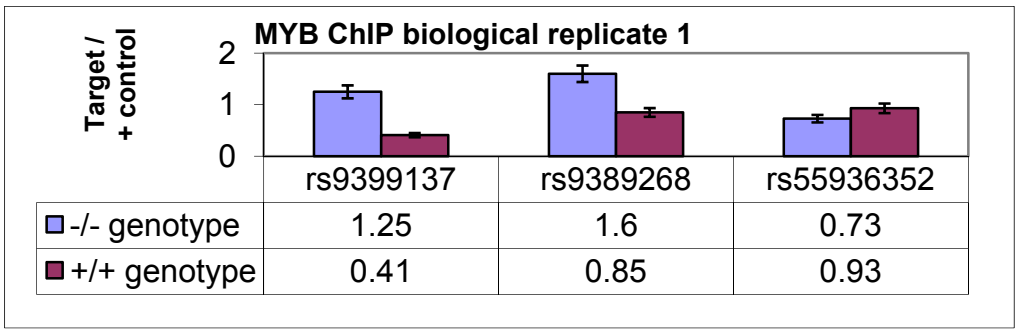
To ensure that we harvest the erythroid progenitor cells from the +/+ and -/- individuals at equivalent stages of the *MYB* expression, we initially obtained a profile of *MYB* expression in cells from the individuals of both genotypes. *MYB* expression profiles suggested that day 4 would be optimal (peak of MYB expression and there is no significant difference between individuals of different genotypes in terms of MYB expression). As validation, the cells were also closely monitored by FACS analysis as described in Chapter 3. When performing ChIP on erythroid progenitor cells at similar

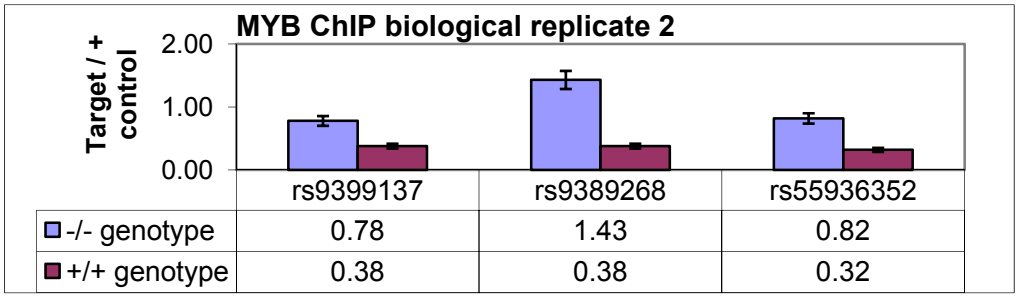
stages of expression, we ensure any difference observed in MYB binding is due to the alteration of binding by the SNPs and not because of differential expression of MYB.

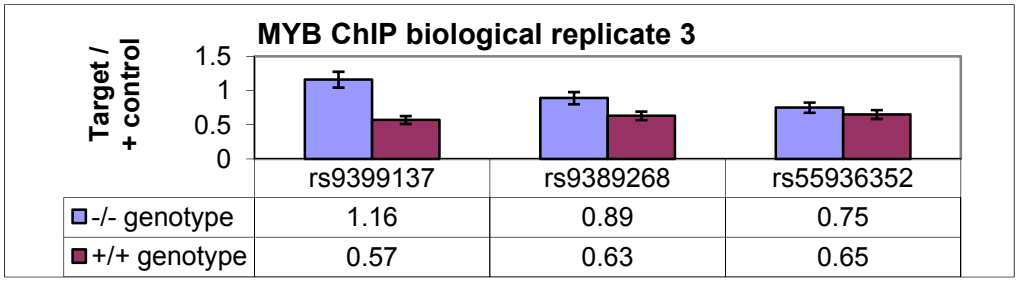
SYBR-Green RT-PCR was performed with material from six erythroid cultures (three HPFH -/- individuals and two HPFH +/+ individuals). Two technical experiments for each culture were performed except for the last pair because of limited material. Enrichments relative to IgG (rabbit immunoglobin G) were calculated to confirm quantitative MYB binding to targets. Results are normalized to positive control CTSG enabling us to cross compare MYB binding to the same target in different individuals.

Our results suggested differential binding of MYB around SNPs rs9399137 and rs9389268, but failed to suggest any differential binding around SNP rs55936352. The results are consistent with the same binding pattern observed in all three biological replicates (Figure 6.12). SNPs rs9399137 and rs9389268 coincide with DNase hypersensitive sites (HS2) and the peak sites of theintergenic transcripts in erythroid progenitors.

A







B

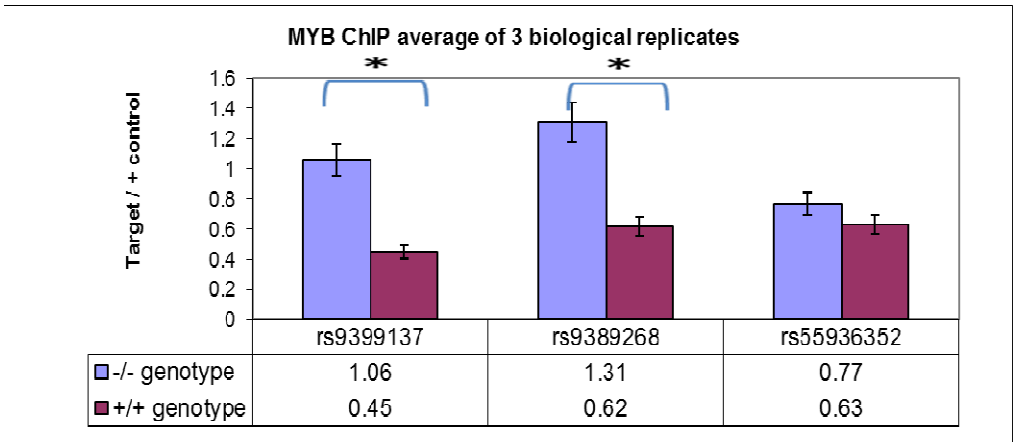


Figure6.13: Differential MYB binding at the 6q23 locus.
A: Binding of MYB was assessed by SYBR-Green RT-PCR on MYB ChIPped chromatin prepared from HEPs from three paired individuals, HPFH +/+ vs. HPFH -/- . Enrichments relative to IgG (rabbit immunoglobin G) were calculated to confirm quantitative MYB binding to targets. Results are normalized to positive control CTSG to enable cross comparison of MYB binding to the same targets in different individuals. B: Mean of MYB binding of 3 replicates. T-test was performed (n: 3) and suggested that there is a significant difference between two groups around rs399137 and rs9389268 (p: 0, 02/ 0, 05).

6.2.3.4 Investigation of allele specific binding of MYB to high scoring SNPs at the 6q23 locus inHPFH +/- erythroid chromatin

Bioinformatics analysis revealed SNP rs9389268 in close proximity to a MYB binding motif. ChIP-qPCR experiments performed with HPFH +/+ and HPFH -/- erythroid chromatin also suggested differential MYB binding around SNP rs9389268. To further investigate and validate differential binding of MYB to the minor allele (+) and major allele (-), we selected individuals that are heterozygous for this key SNP for allele specific ChIP. Healthy unrelated individuals were recruited and genotyped for the SNP of interest (rs9389268). Cell cultures were performed with erythroid progenitor cells isolated from peripheral blood of individuals who were heterozygous for the candidate SNPs.

Erythroid progenitor cells were harvested and cross-linked at the stage of peak MYB expression i.e. day 4 of phase II erythroid cultures. Cross-linked material was immunoprecipitated with the same MYB (1:1 Millipore) antibody and chromatin was eluted as previously described. Eluted chromatin was tested for enrichment by SYBR green RT-PCR. Enrichment of positive control CTSG and targetof interest (rs9389268) was calculated (Figure 6.13).

The results suggest weaker binding of MYB in chromatin of HPFH +/- individuals to minor allele G that is associated with elevated HbF. These results are consistent with the results observed with experiments performed with HPFH +/+ and HPFH -/- individuals. The main difference and the main advantage of this approach over ChIP-qPCR experiments is that the differential binding is internally controlled and the results are much more robust compared to previous ChIP-qPCR experiments.

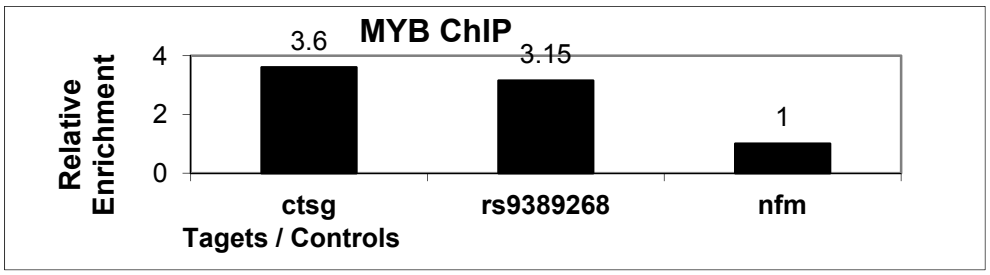
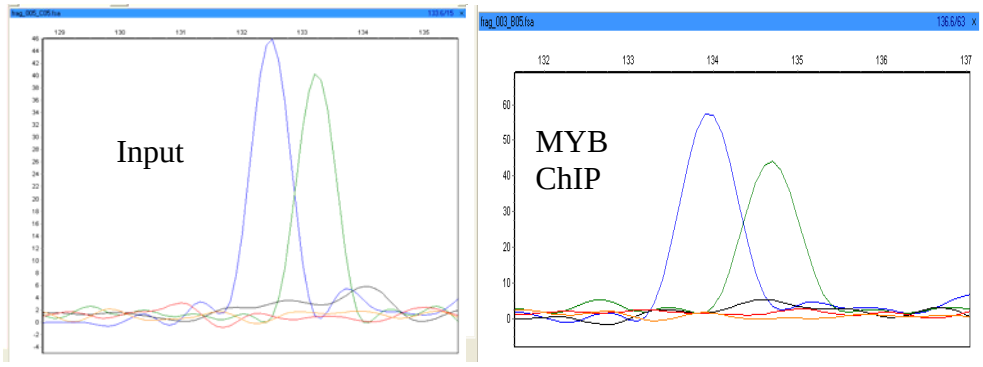


Figure6.14: SYBR-Green RT-PCR was performed to confirm MYB enrichment of positive control CTSG and target rs9389238.
Shown as fold enrichment, normalized to *NFM* as a housekeeper gene and to IgG (ChIP with mock antibody). Enrichment results are Mean values of 3 biological replicates.

A



B

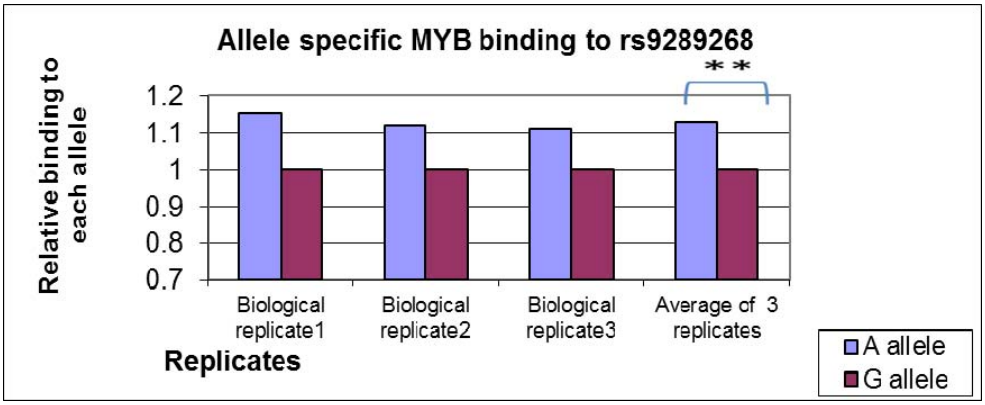


Figure6.15: Allele specific MYB binding to rs9389268.
A: Snapshot was performed with chromatin immunoprecipitated with MYB (clone 1:1 Millipore) and input. Ratio of MYB binding to major allele A and minor allele G was calculated by using gene marker program which provides the peak height of both alleles. Calculation were done with both chip material and input for each individual. B: Experiments were performed with three biological samples that were heterozygotes for the SNPs of interest .A to G ratio of chip material was calculated and normalized to input. T-test was performed (n: 3) and suggested that there is significant difference between two alleles (p: 0.004).

6.2.4 KLF1 binding at chromosome 6q QTL

In-silico experiments suggested the presence of KLF1 binding sites at chromosome 6q and this was validated by ChIP experiments. Colleague PhD student Kiran Jawaid performed ChIP-chip with KLF1 but failed to suggest any strong KLF1 sites at HMIP block two. I applied ChIP-qPCR to investigate the potential binding sites.

6.2.4.1 Control test for KLF1 enrichment

Following confirmation of the chromatin size, we proceeded to immunoprecipitation using an anti KLF1 antibody (provided kindly by Prof. Sjaak Philipsen). After reverse cross-linking and purification of the ChIP DNA, a small amount was tested for enrichment by SYBR green RT-PCR. A site in the promoter region of beta globin gene (*HBB*) is a known KLF1 binding site and was used as positive control in KLF1 ChIP experiments. This step is a check point to confirm if ChIP with the target antibody has been successful.

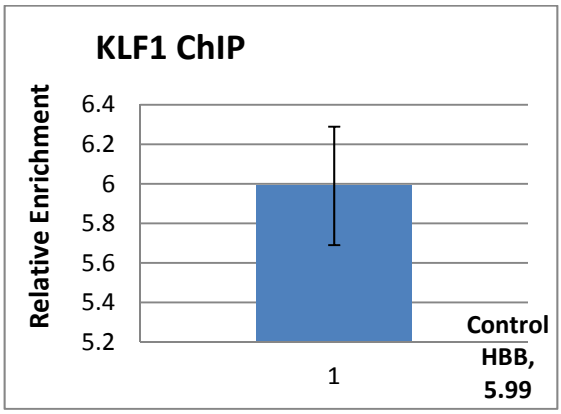


Figure6.16: SYBR-Green RT-PCR for HBB enrichment was performed to confirm KLF1 ChIP (using antibody KLF provided by Prof. Sjaak Philipsen).
Shown as fold enrichment, normalized to *NFM* as a housekeeper gene and IgG (ChIP with mock antibody). Enrichment results are Mean values of 5 biological replicates +/- standard deviation.

6.2.4.2 Quantitative KLF1 binding at the 6q23 locus (HMIP)

Previous *in-silico* experiments demonstrated the presence of KLF1 binding motifs at specific positions at the 6q23 locus (*HMIP*) (See Chapter 3). The presence of some of the KLF1 binding sites were validated by EMSA experiments (Chapter 4). Although EMSA experiments are informative the results were obtained *in vitro* from k562 cell lines. We wanted to validate KLF1 binding to predicted sites 6q23 locus in primary human erythroid progenitor cells by ChIP-qPCR.

Primers were designed for the sites that were strong candidates based on *in-silico* experiments. Experiments were performed to optimise conditions for qPCR on KLF1 ChIP material from healthy individuals. Mock IgG method was used in calculation of the enrichment on target sites. This method depends on normalization with a mock IgG: Fold enrichment = $2^{(ct\ IgG-ct\ ab)}$ (Figure 6.16).

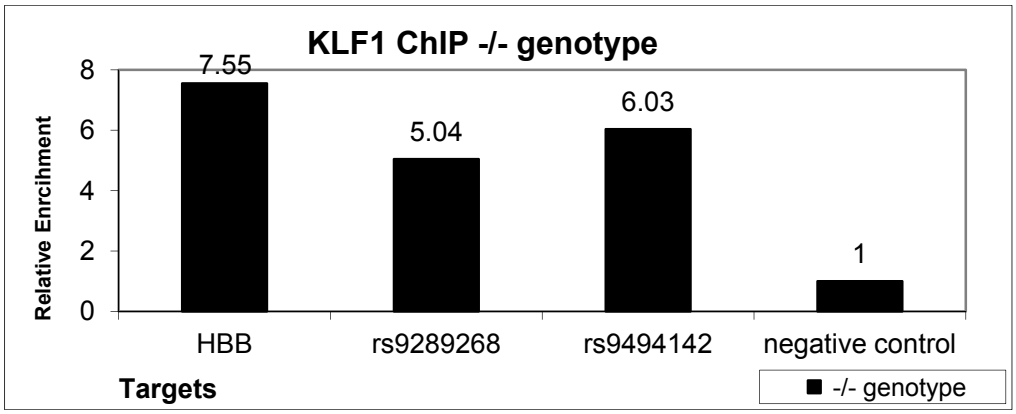


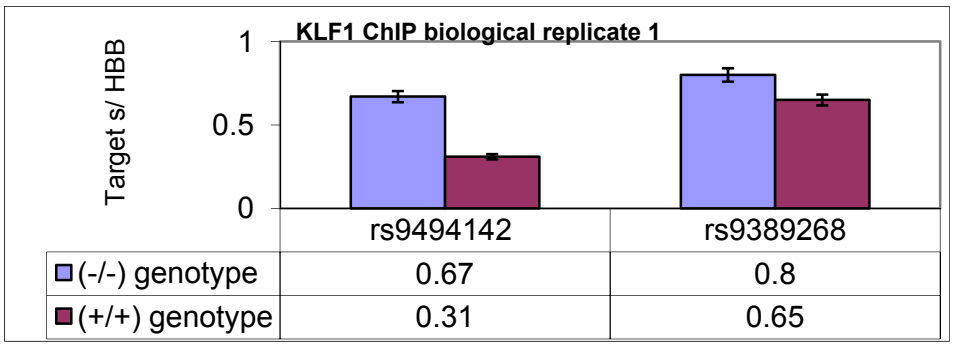
Figure6.17: KLF1 Binding at chromosome 6q: SYBR-Green RT-PCR was performed.
Enrichments relative to IgG (rabbit immunoglobin G) were calculated to confirm quantitative KLF1 binding to targets. HBB was used as positive control and *NFM* was used as a negative control.

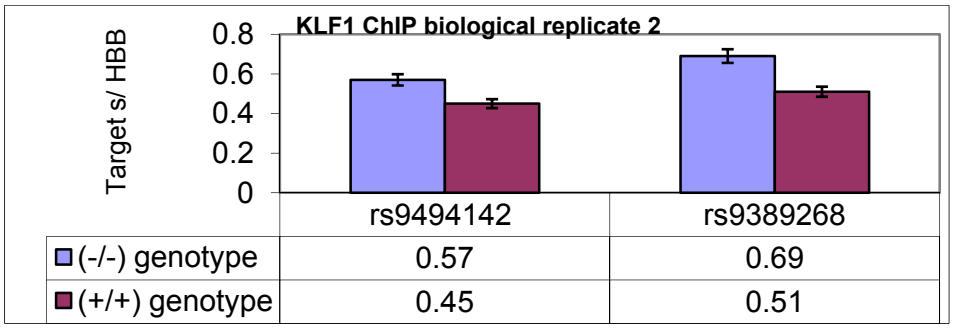
Strong KLF1 binding was observed around SNPs rs9389268 and rs9494142 (Figure 6.17). We investigated different KLF1 antibodies and the best enrichment was observed with KLF1 antibody provided by Prof. Sjaak Philipsen.

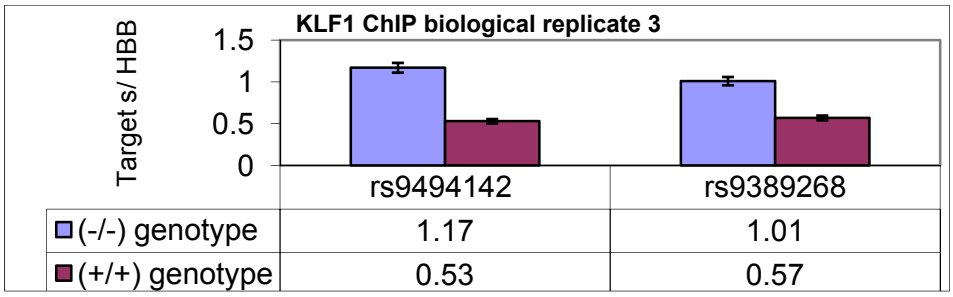
6.2.4.3 Investigation of differential KLF1 binding at the 6q23 locus using ChIP-qPCR in HPFH +/+ vs HPFH -/- erythroid chromatin

Erythroid progenitor cells were obtained from primary cultures of whole blood taken from three paired sets of HPFH +/+ and HPFH -/- individuals. The +/+ individuals were members of the same family in which the QTL was first identified. One +/+ individual donated blood twice. ChIP-qPCR was carried out as previously described using the same anti-KLF1 antibody. To ensure that we harvest the erythroid progenitor cells from the HPFH +/+ and -/- individuals at equivalent stages of *KLF1* expression, we initially obtained a profile of *KLF1* expression in cells from individuals of both genotypes. *KLF1* expression profiles suggested that phase II day 7 would be optimal, when KLF1 is at the peak of its expression and there is no significant difference between individuals of different genotypes in terms of KLF1 expression. As validation, the cells were also closely monitored by FACS analysis as described in Chapter 3. Hence chromatin obtained at phase II day 7 of erythroid progenitor cells was used for both GATA-1 and KLF1 ChIP-qPCR experiments. In performing ChIP on erythroid progenitor cells at similar stages of expression, we ensure any difference observed in KLF1 binding is due to the alteration of binding by the SNPs and not because of differential expression of KLF1.SYBR-Green RT-PCR was performed with material from six erythroid cultures (three HPFH -/- individuals and two HPFH +/+ individuals). Two technical experiments for each culture was performed except the last paired because of limited material. Enrichments relative to IgG (rabbit immunoglobin G) were calculated to confirm quantitative KLF1 binding to targets. Results are normalized to positive control HBB enabling us to cross compare KLF1 binding to the same target in different individuals. Results suggested the differential binding of KLF1 around SNPs rs9389268 and rs9494142. The results are consistent with the same binding pattern observed in all three biological replicates (Figure 6.17). SNPs rs9389268 and rs9494142 coincide with DNase hypersensitive sites (HS2 and HS3) and the peak sites of intergenic transcripts.

A







B

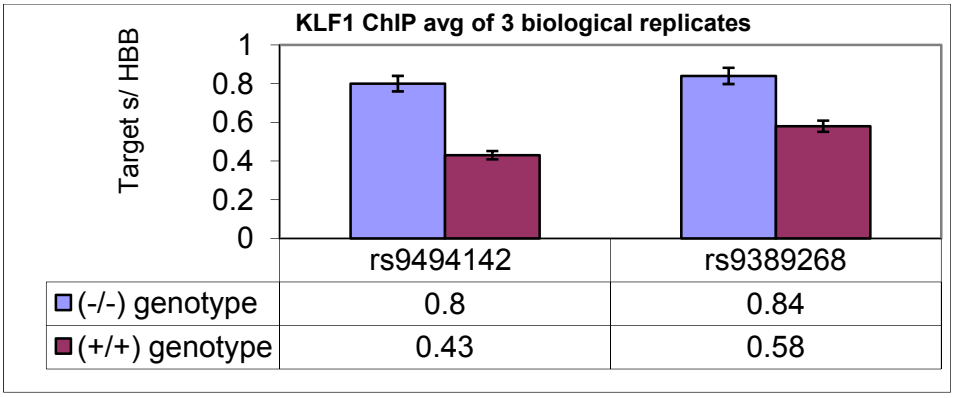


Figure6.18: Differential KLF1 binding at the 6q23 locus.
A: Binding of KLF1 was assessed by SYBR-Green RT-PCR on KLF1immunoprecipitated chromatin prepared from HEPs from three paired individuals, HPFH +/+ vs. -/- . Enrichments relative to IgG (rabbit immunoglobin G) were calculated to confirm quantitative KLF1 binding to targets. Results are normalized to positive control HBB to enable cross comparison of KLF1 binding to the same targets in different individuals. B: Mean of KLF1 binding of replicates. T-test was performed (n: 3) and suggested that there is no significant difference between two groups.

6.2 Discussion

A QTL for HbF levels and F cells was localized to chromosome 6 by genetic association studies. High-resolution genotyping and further analysis refined the location to an intergenic region between the *HBS1L* and *MYB* genes (*HMIP)*. These SNPs lie in 3 linkage disequilibrium blocks, *HMIP 1-3.*

HMIP2 accounts for the strongest effect on FC variance at this locus. Using chromatin immunoprecipitation experiments, our group showed that this area has significant histone acetylation and RNA polymerase II activity. Several strong GATA-1 binding sites are concentrated within *HMIP2,* three of which coincide with DNase I hypersensitive sites. Such patterns of RNA polymerase II binding, histone acetylation and GATA-1 binding around DNase I hypersensitive sites is seen in known erythroid control regions, suggesting that this intergenic space may contain major distal regulatory elements (Wahlberg et al 2009).

GATA-1 is a well-known erythroid transcription factor, which acts as an activator of the globin genes. Its role as a transcription factor is critical to erythropoiesis, which made it an important factor to investigate at the 6q QTL. Investigating GATA-1 binding at this locus would lead to a better understanding of how this QTL is regulated and how it may be directly or indirectly influence HbF levels.

In this thesis, I report the GATA-1 binding profile of the previously detected sites quantitatively. The results confirm the binding of GATA-1 to all sites that were previously detected by ChIP-chip. GATA-1 ChIP using erythroid chromatin from three paired sets of erythroid progenitor cells from HPFH +/+ and -/- individuals confirmed the differential GATA-1 binding patterns between the HPFH +/+ and HPFH -/- genotypes previously shown by fellow colleague Kiran Jawaid.

In the previous ChIP-chip experiments, GATA-1 bound strongly a site designated as GATA-1 peak 4 in the individuals with -/- genotype, however no GATA-1 binding signals were detected in the individual with +/+ genotype. Although these binding signals were confirmed by two technical repeats on one biological sample for each individual (Figure 6.5), these results are not quantitative and should be validated by qPCR. To confirm the differential binding quantitatively, quantitative Real Time PCR (qPCR) on GATA-1 immunoprecipitated chromatin was performed on the previously detected sites. Results suggested that GATA-1 binding is altered in the region encompassing the SNP rs94944142. The GATA-1 binding is weaker in peak four in individuals homozygous for SNPs associated with elevated HbF (Figure 6.6). ChIP-qPCR results also suggested differential binding of GATA-1 to another key area: GATA-1 peak 3. This region also encompasses key variant i.e. (SNP rs66650371) that is associated with elevated HbF and coincides with conserved GATA-1 sites and DNase hypersensitive sites.

ChIP-qPCR results suggest differential GATA-1 binding to key regions, which also encompass variants that are associated with elevated HbF (Figure 6.6). As further validation, I have performed allele specific ChIP experiments on erythroid chromatin from individuals heterozygous for these SNPs. In both cases, the results confirmed differential GATA-1 binding to specific SNPs (rs9494142 and rs66650371) and additionally, weaker binding of GATA-1 to the minor (+) allele that is genetically associated with elevated HbF levels is demonstrated. These results are consistent with previous ChIP-qPCR results and increase the likelihood of these SNPs being the causative variants in the region. Here, we provide functional confirmation for the *in-silico* transcription factor predictions. The differential GATA-1 binding at this site supported the recent study by Farrell et al who identified enhancement activity of a region in the chromosome 6q which includes GATA-1 binding motifs.

Differential binding of GATA-1 to our target sites may lead to a decrease in the enhancement activity around those target sites and to differential expression of flanking genes (Farrell et al 2011).

Another recent study by Stadhouders et al demonstrated that a transcription factor complex including LDB1, GATA-1, TAL1, ETO2 and KLF1 is also present at the chromosome 6q. The group proposed a model for *Myb* activation by distal enhancers which are dynamically bound by KLF1 and the GATA1/TAL1/LDB1 complex, primarily functioning as a transcription elongation element through chromatin looping. Previous studies showed that this complex is required to maintain chromatin looping to form an active chromatin hub (ACH) at the β-globin locus to initiate transcription (Tolhuis et al2002, Song et al2007).

The study by Stadhouders et al also demonstrated that in the proliferating cells expressing *Myb*, enhancers within the *Myb-Hbs1l* intergenic region were shown to form an active chromatin hub (ACH). The *Myb* promoter and first intron are included within the ACH and it is lost when cells terminally differentiate, concomitant with the downregulation of *Myb* and decreased binding of TF complexes at the distal enhancers. These LDB1 complexes were found to bind at sites in the *HBS1L-MYB* intergenic region in MEL cells and primary mouse erythroid progenitors . These sites coincide with GATA-1 peak three and four in our primary human erythroid progenitors. Further experiments showed characteristic features supporting enhancer activity (including the presence of histone acetyl transferase p300) and led to the conclusion that these binding sites contain active regulatory elements in erythroid progenitors(Stadhouders et al 2012).

This study provided further understanding about the function of the *HBS1L-MYB* intergenic region and supports the idea that regions around GATA-1 peak three and four located at -84kb and -71kb away from the MYB transcription start site are key regions of *HMIP* block two.

The altered KLF1 binding observed around SNP rs9494142 can be due to the presence of KLF1 and GATA-1 in the same complex. Tallack et al demonstrated that the vast majority of KLF1 binding sites are located at large genomic distances from known transcription start sites and are co-occupied by GATA-1 (Tallack et al 2010).GATA-1 binding could be necessary to initiate the binding of the complex and the altered binding of GATA-1 may lead to altered binding of other transcription factors in the complex. Tallack et al also suggests that KLF1 is not only capable of, but necessary for, long-distance gene regulation within erythroid cells which could explain the differential regulation of *MYB* and/or *HBS1L* genes due to disruption of KLF1 binding.

Differential binding of KLF1 around SNP rs9389268 was also suggested from the work carried out in this thesis. EMSA results suggested a potential complex formation between KLF1 and MYB and ChIP results suggested altered binding of MYB and KLF1 factors around SNP rs9389268. Co-occupation of a region could suggest a complex formation between MYB and KLF1 but further studies are necessary to support this suggestion.

MYB and KLF1 transcription factors are both crucial regulators of erythropoiesis and several studies show that there is a correlation between the expression profiles of these factors. Bianchi et al suggested that MYB transcription factor promotes erythropoiesis by transactivation of human KLF1, likely within the upstream enhancer region. ChIP and enhancer assays by the same group demonstrated that MYB binds to *KLF1* promoter.

Another study by Suzuki et al suggested *KLF1* expression is downregulated in Myb knockdown mouse (Bianchi et al 2010, Suzuki et al 2013). The importance of KLF1 binding in long-distance regulation is already stated and in this case, it is possible that MYB with KLF1 regulate the flanking gene *MYB* which would explain the difference in

MYB expression observed in individuals with different genotypes in this region (Jiang et al 2007).

Another possibility is the auto regulation of *MYB* gene by its own transcription factor and the altered binding of MYB can lead to down regulation of *MYB* due to altered activation by the transcription factor. *MYB* gene expression should be strictly regulated because of its key roles during erythropoiesis. A feedback mechanism of *MYB* would prevent overexpression of *MYB* and a positive autoregulation would enable quick response in expression when MYB transcription factor is needed. *S*tudies showed that key transcription factors like GATA-1 bind to their own promoters to regulate the expression. Nishikawa*et al* showed, in transgenic zebra fish, that Gata1 positively auto-regulates its own expression. The group demonstrated that mutations of Gata1 that decreased self-association greatly reduced the transactivation activity of Gata1. A similar mechanism is possible for *MYB* regulation. A study suggested that MYB transcription factor binds to 5' flanking region of the *MYB* gene positively autoregulating the expression of the gene in order to maintain high levels of MYB protein in cells that are proliferating (Nicolaides et al 1991). Hence, differential binding of the MYB transcription factor to candidate SNPs may lead to differential expression of *MYB* which clearly influences HbF level and provide insights on the causal SNPs.

7 ALLELE SPECIFIC EXPRESSION OF HMIP FLANKING GENES

7.1 Introduction

7.1.1 Chromosome 6q QTL

A quantitative trait locus (QTL) controlling HbF levels was mapped to a 1.5 Mbregion on chromosome 6q23 in an Asian-Indian kindred with beta thalassaemia and heterocellular hereditary persistence of fetal hemoglobin (HPFH). Annotation of this region showed that it contained five protein coding genes, *ALDH8A1*, *HBS1L*, *MYB*, *AHI1*, and *PDE7B*of which two *(MYB* and*HBS1L)*were expressed in erythroid tissue . The region also contained 5 pseudo genes and 4 non-coding transcripts(Close et al 2004, Jiang et al 2006).

Further experiments focused on the 6q23 transcriptome to investigate if 5 candidate genes in the region are expressed in the erythroid progenitor cells and if there is differential expression of these genes between high F and low F individuals. A panel of high F and low F individuals were recruited for primary erythroid cultures and expression profiles of the 5 genes at different stages of phase II were obtained. The results demonstrated that only*MYB* and *HBS1L* genes are expressed in erythroid progenitor cells. Further analysis of the results suggested that *MYB* and *HBS1L* are quantitative trait genes, with variable expression in the erythroid progenitor cells of healthy adults. Analysis of erythroid cultures from a number of individuals has revealed a clear correlation of elevated peripheral blood HbF with low levels of *MYB* and *HBS1L* expression in erythroid cultures. Furthermore, over expression of MYB in K562 cells inhibited γ-globin gene expression when compared with the parental cell line. However, overexpression of HBS1L had no effect on γ-globin gene expression (Jiang et al 2006).

High resolution genetic mapping subsequently refined the 1.5 Mbintervals to an intergenic region of 79 kb located between the *HBS1L* and *MYB* genes. It would seem that the QTL which reside in the intergenic region regulate HbF indirectly via control of the flanking genes. Our group then showed the presence of regulatory elements in the *HBS1L-MYB* intergenic region (*HMIP*) evidenced by GATA-1 binding coinciding with strong histone acetylation, RNA polymerase II activity, and erythroid specific DNase I hypersensitive sites(Wahlberg et al 2009). The presence of enhancer elements were supported by mouse studies in which interaction of the homologous region with the promoter and first intron of the flanking *Myb* gene via chromatin looping was demonstrated (Stadhouders et al 2012).

The next step towards discovery of the causative SNPs involved comparing the detailed sequence of the region between the high F and low F individuals. Resequencing of two key individuals (HPFH +/+ vs. HPFH -/-) from the family that led to discovery of the 6q QTL revealed new SNPs in the *HBS1L-MYB* intergenic region. Genetic studies suggested some of the key SNPs were strongly associated with elevated HbF levels. The key SNPs were located in two regions that coincided with transcription factor binding sites. Importance of these sites was supported by EMSA and ChIP experiments that suggested these variants alter transcription factor binding.

The next step was to associate these variants with the expression of the flanking genes i.e. *MYB* and/or *HBS1L.* We hypothesise that these SNPs affect expression of the flanking *MYB* gene but not that ofHBS1L .To demonstrate unequal expression, we recruited individuals who are heterozygous for both the intergenic SNPs and heterozygous for the expression markers in the flanking genes. Individuals who are homozygous (HPFH +/+ and HPFH -/-) for intergenic SNPs but heterozygous for expression markers in the flanking genes were also recruited as controls.

7.1.1.1 MYB

The *MYB* gene (also known as *c-Myb*) was first identified as the cellular homologue of *v-Myb*, an oncogene found in two avian retroviruses that induce leukemia.
The *MYB* gene is highly conserved through evolution (Lipsick, 1996); the major product of the human gene is a 72kDa nuclear protein of 640 amino acids. The protein is expressed in most haematopoietic tissues and is highly expressed in immature haematopoietic cells (Gonda et al 1982, Westin et al 1982) (Figure 7.1).

The MYB protein (72 kDa) has an isoform of 89 kDa (MYB-9A) caused by a splice variant which leads to an insertion of 363 base pairs (bp) between exon 9 and 10 (exon 9A). The p89 form of MYB displays higher transactivation activity, however little is known about its functional relevance. Jiang et al confirmed that the pattern of *MYB-9A* expression was similar to that of the total *MYB* expression in erythroid progenitors.

The MYB transcription factor plays crucial roles in regulation of proliferation and cell lineage commitment of erythroid progenitor cells. This is supported by the fact that mice with myb knock-down will die at embryo stage due to the lack of erythroid and myeloid development.MYB is required for maintenance of proliferation by exerting a direct role in cell cycle control. This transcription factor is highly expressed in immature haematopoietic cells and is down-regulated as the cells differentiate. Expression of *MYB* blocks differentiation in neuroblastoma cell lines, suggesting down-regulation of *MYB* is also required for terminal differentiation (Westin et al 1982, Oh et al 1999).
In addition to controlling erythropoiesis, the MYB transcription factor has several other essential roles throughout the different stages of haematopoiesis as well as being involved in leukaemogenesis (Greig et al 2008). Despite its importance in many biological systems, the regulatory mechanisms controlling MYB expression are still not clear. Various factors including GATA-1 (Bartunek et al 2003), Ets-1 (Sullivan et al 1997) have been suggested to regulate MYB expression. Other studies have shown

that *MYB* introns contain regulatory sequences which play an important part in MYB control and are represented by an attenuation block in the first intron (Bender et al 1987) and enhancer elements in introns 1 and 4 (Dooley et al 1996;Seib et al 1994).

Studies have demonstrated that MYB is a key target of the microRNA (miRNA) miR-150 (Xiao et al 2007)and that one pathway of MYB regulation is through the two conserved miR-150 binding sites in the 3' UTR of MYB mRNA. miR-150 repression of MYB in $CD34^{+}$ human bone marrow cells not only supported *MYB*'s key role in erythroid and megakaryocytic differentiation, but also suggested that modulations of its level is critical to its role in haematopoiesis (Lu et al 2008).

Jiang et al demonstrated that *MYB* expression in erythroid progenitor cells of healthy adults is variable and that individuals with higher HbF levels have relatively lower MYB levels. It is likely that *MYB* is a quantitative trait gene and that level of the MYB product is controlled by *cis*-regulatory and *trans*-acting factors.*MYB* thus appear to be the prime candidate for our allele specific expression study since altered expression of *MYB* clearly influences HbF levels.

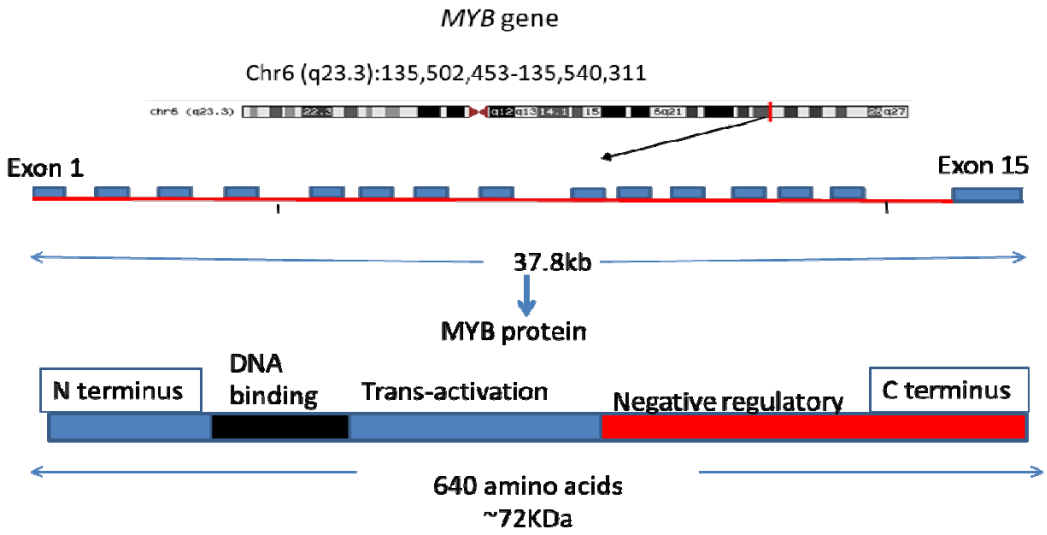


Figure 7.1: *MYB* gene and MYB protein structure.
MYB gene spans a region of 37.8 kb and consists of 15 exons coding for a protein that is formed of 640 amino acids. The MYB transcription factor has a DNA binding domain close to N terminus, a central trans-activation domain and a negative regulatory domain.

7.1.1.2 *HBS1L*

The Hsp70 subfamily B suppressor 1-like gene *(HBS1L)* is the human ortholog of the *Sacharomyces cerevisiae HBS1* gene. It encodes a GTP-binding protein which shares structural features with elongation factor eEF-1A and release factor 3 (eRF3) families (Wallrapp et al 1998).Multiple alternatively spliced transcript variants encoding different protein isoforms have been found for this gene but they are not well characterized. *HBS1L* is expressed in multiple tissues with the highest expression in heart and skeletal muscle.

Recent genome-wide datasets generated by the ENCODE consortium shows that *HBS1L* was expressed at similar levels in all cell types and was not restricted to hematopoietic cells. Analysis of the results also reveals that there is no correlation between the expression of *HBS1L* and the HBS1L-MYB intergenic regulatory activity. Its role, if any, and to what extent in the regulation of HbF, is yet to be determined. QTL mapping together with previous genome-wide association study involving approximately 110,000 gene-based SNPs in mild and severe beta-0-thalassemia/Hb E patients revealed SNPs in HBS1L significantly associated with Hb F levels and severity of the disease. One study focused on a SNP located in *HBS1L* exon 1 that was suggested to be associated with severity in beta thalassemia/Hb E patients
The variant was a C32T polymorphism that was identified in a potential E-box binding site of HBS1L exon 1 and it was speculated that disruption of the binding site can lead to the elevated HbF levels. The SNP was genotyped in 455 beta thalassemia/Hb E patients and results suggested that C allele is associated with elevated HbF levels (Pandit et al 2008).

7.1.2 Allele specific expression

Various studies have indicated that expressions of a large proportion of human genes are subject to variable cis-acting influences (Bray et al 2003). Cis-acting variants

affecting gene expression may be located in the transcribed sequence itself, promoters, intronic sequence or in distant regulatory elements (Levine and Tjian, 2003). In the absence of cis-acting influences differentially affecting the transcription or stability of the mRNA from each copy of an autosomal gene, both copies of that gene will be equally expressed. Identification of unequal allelic expression in an individual heterozygous for any cis-acting polymorphism affecting transcription suggests that the gene is subject to cis control.

Assays of relative allelic expression is an effective approach for exploring the influence of cis-acting regulatory SNPs. Assays of relative allelic expression expose only the variance in gene expression resulting from cis-acting effects since it is an intra-individual approach. Each allele acts as an internal control for the other, allowing detection of genuine cis-acting effects, while controlling for the trans-acting environment. Thus assay of relative allelic expression is a powerful approach for examining the influence of specific haplotypes on expression of the target gene (Bray et al 2005). Our allelic expression experiments have been performed using Snapshot, a technique with quantitative properties with the aid of Nicholas Bray and Matthew Hill (Institute of Psychiatry, Kings College London).

7.2 Results

We have performed experiments following two different approaches to associate the intergenic variants with the expression of the flanking *MYB* and *HBS1L* genes. Initially, we have compared *MYB* and *HBS1L* expression profiles in HPFH +/+ vs. HPFH -/- individuals. Next, we have performed allele specific expression experiments with HPFH +/- individuals to confirm the association of intergenic variants in *HMIP-2* with the expression of the flanking *MYB* and *HBS1L* genes intraindividually.

7.2.1 Expression profiles of *MYB* and *HBS1L* within erythroid progenitor cells

Previous studies in our group have shown that *MYB* and *HBS1L* are quantitative trait genes with variable expression during erythropoiesis. Expression profiles of each gene at different stages of erythropoiesis in primary cultures of individuals with different HbF levels and different genotypes at the 6q intergenic interval would be most informative. The profiles would also provide information for harvesting cells at the optimum stage for allele specific gene expression experiments

We constructed expression profiles of the *MYB* and *HBS1L* genes from erythroid cells at sequential days of Phase II. Ct values were normalised to the house-keeping*HPRT* gene. *MYB* expression peaks at early stages of phase II day (2-4), then gradually decreases as the cells differentiate (Figure 7.1 A). In contrast, *HBS1L* expression fluctuates during phase II with no clear expression pattern during differentiation (Figure 7.1 B). From the expression profile it would appear that phase II day 3 is an ideal stage for harvest since both genes are at the peak of their expression. This stage coincides with the highly proliferative stage when the cells are mainly pro-erythroblasts (Chapter 3).

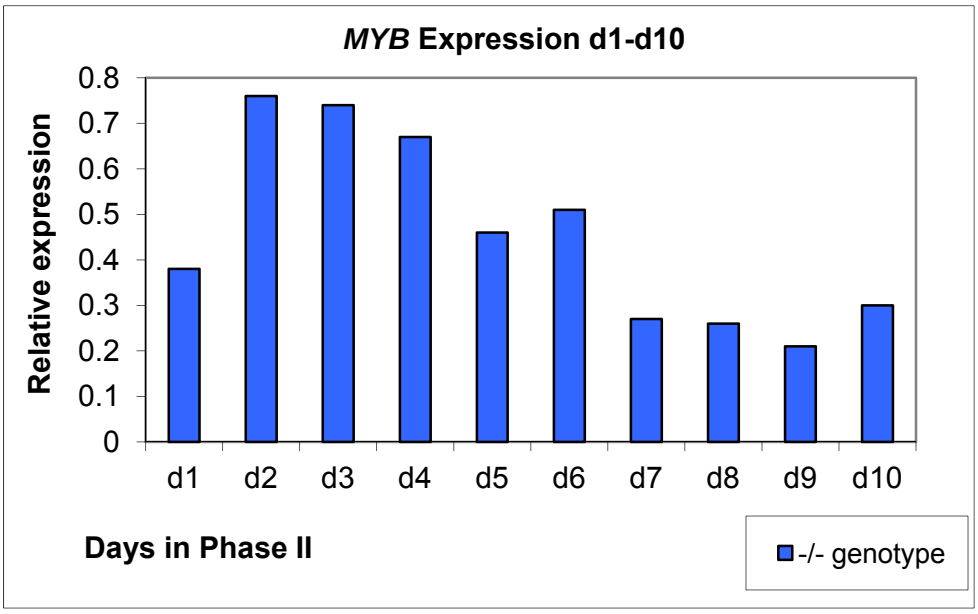
7.2.2 Comparison of *MYB* and *HBS1L* expression profiles in HPFH +/+ vs. HPFH -/- individuals

The study by Jiang et al was performed using erythroid progenitor cells from individuals who were phenotypically selected as high HbF or low HbF. Here, we obtained primary erythroid cells from high HbF individuals who were also homozygous for all the variants in the HBS1L- MYB intergenic region that are associated with elevated HbF levels, and low HbF individuals who are also homozygous for the absence of variants associated with elevated HbF levels. These individuals were referred to as HPFH +/+ and HPFH -/-, respectively. Cells cultured from individuals of both genotypic haplotypes shared a similar expression pattern with peak of expression for *MYB* gene at day 2-4 of phase II and variable expression for *HBS1L*.

MYB expression peaks at relatively early stages of phase II and then gradually decreases in both genotypes (Figure 7.3). Cells cultured from HPFH +/+ individuals showed lower *MYB* expression levels at later stages (day5 –day 10) of the culture when compared to *MYB* expression levels of erythroid cells cultured from low HbF individuals. The result is more striking when MYB expression is normalized to day 4 (Figure 7.4). The sharpest decrease in *MYB* expression can be observed at day 5 in HPFH +/+ individuals; in contrast, the sharp decrease is observed later at day 7 in HPFH -/- individuals suggesting a difference in the MYB expression pattern between two genotypeslater stages os culture. These are preliminary results and experiments with 20 individuals (3 +/+, 8-/- and 9+/-) are being performed to investigate expression patterns of *MYB* with different genotypes.

Expression pattern of *HBS1L* between individuals of the two different genotypes (HPFH +/+ vs. HPFH -/-) was presented. The results are from a pair of individulas and suggests that a larger sample size for verification of MYB expression with genotype in the HBS1L-MYB is needed.(Figure 7.3).Experiments on erythroid RNA from an additional20 individuals (3 +/+, 8-/- ,and 9+/-) are being performed to investigate expression patterns of *HBS1L*with different genotypes.

A



B

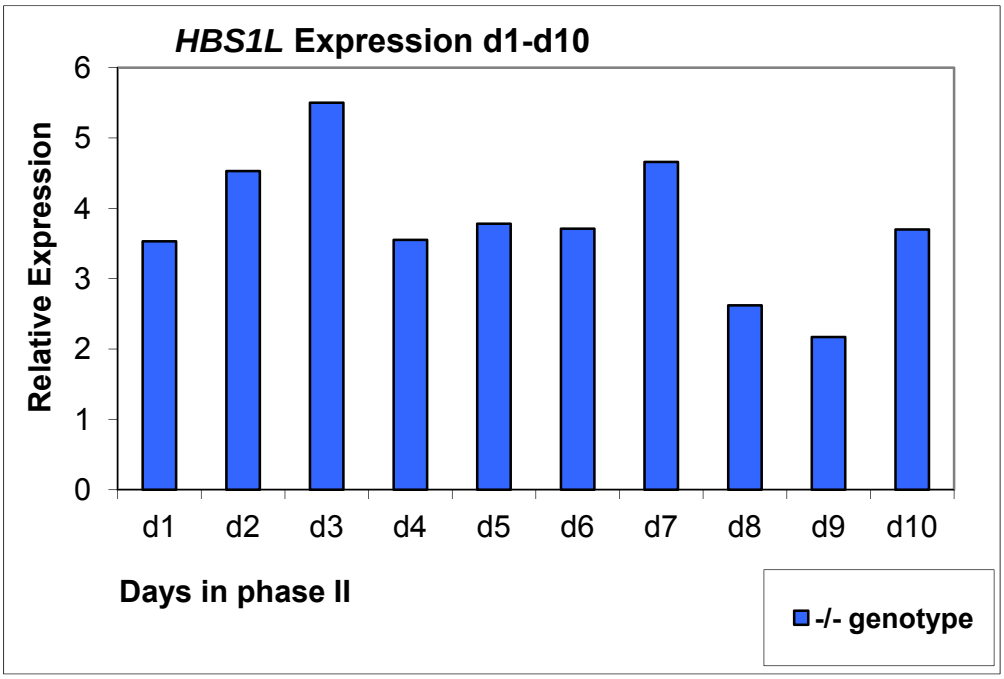
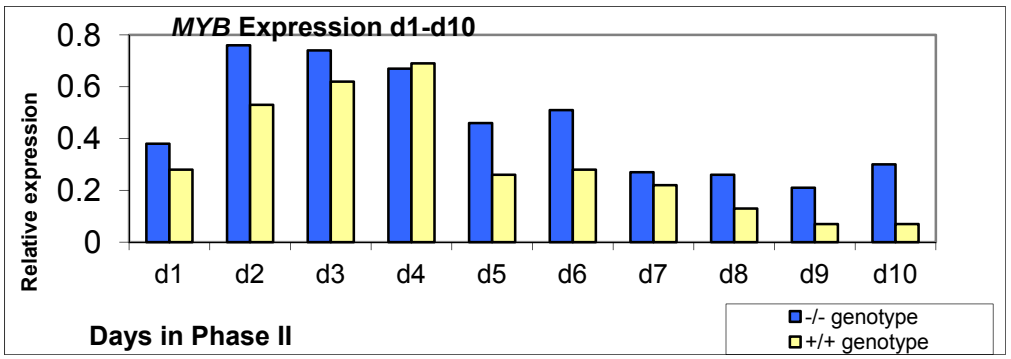
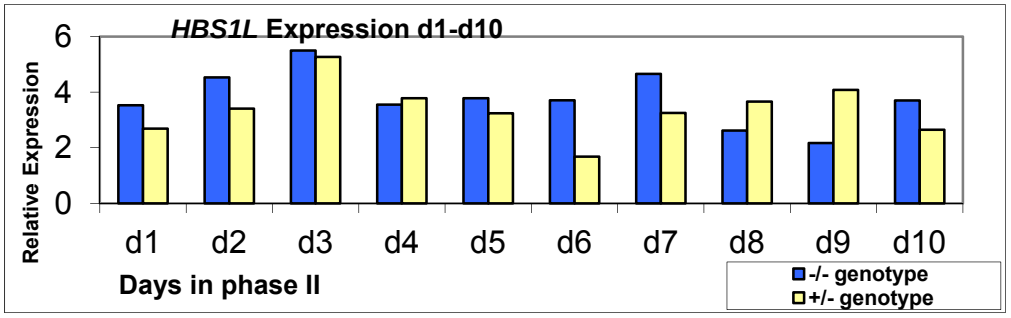


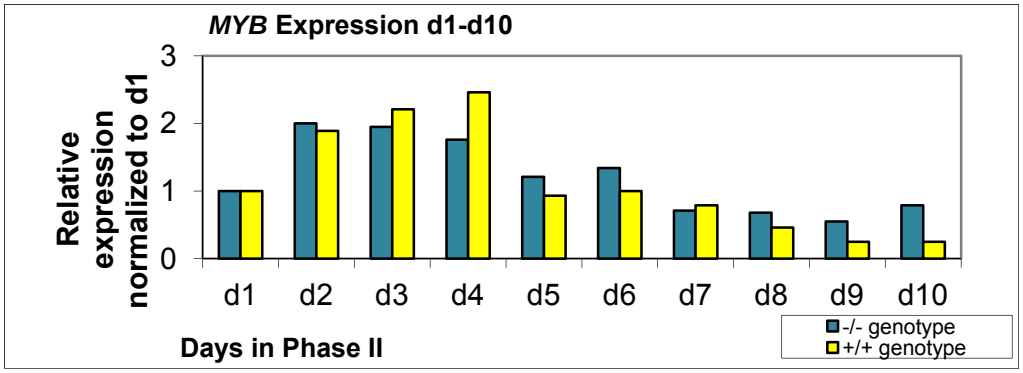
Figure7.2: Relative MYB and HBS1L expression profile from phase II day1-10 HEPs:
Expression profiles of MYB and HBS1L in primary human erythroid progenitor cells taken on sequential days (1 - 10) of Phase II culture. Results are analysed by RT-PCR in duplicate using gene specific primers on erythroid RNA from a single erythroid culture. Individual designated as -/- is homozygous for the absence of HMIP block 2 SNPs associated with elevated HbF. Data were normalized to housekeeping gene HPRT.

A





B



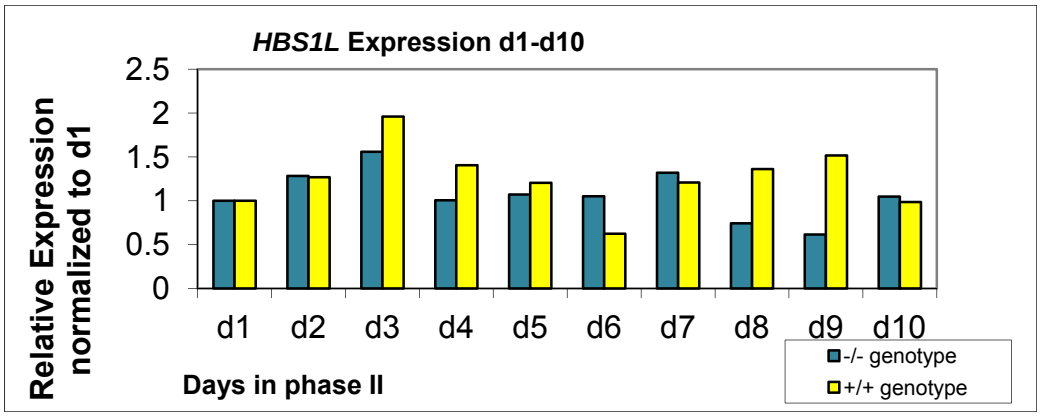


Figure7.3: Relative MYB and HBS1L expression profile from phase II day1-10 HEPs:

Expression profiles of MYB and HBS1L in primary human erythroid progenitor cells taken on sequential days (1 - 10) of Phase II culture. Results are analysed by RT-PCR in duplicate using gene specific primers on erythroid RNA from a single erythroid culture.

Individual designated as -/- is homozygous for the absence of HMIP block 2 SNPs associated with elevated HbF. Individual designated as +/+ is homozygous for the presence of HMIP block 2 SNPs associated with elevated HbF. To cross compare expression profiles of +/+ vs. -/- individuals, expression profiles were normalized to HPRT (Figure 7.3 A) and then normalized to d1 expression (Figure 7.3 B).

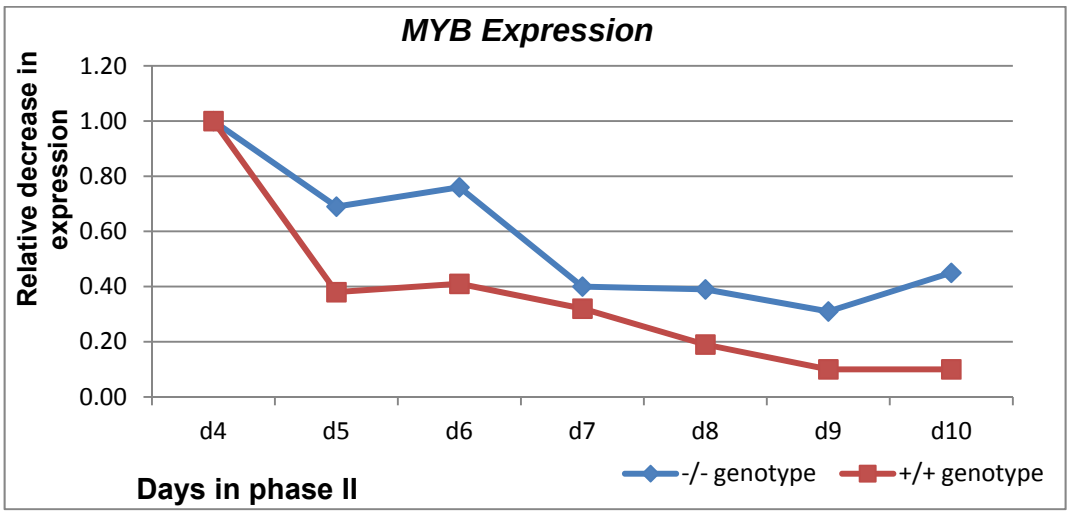


Figure7.4: Relative decrease in *MYB* expression from day 4 to day 10.
To cross compare relative decrease in *MYB* expression with +/+ vs. -/- individuals, expression profiles were further normalized to day 4 expression.

7.2.3 Allele Specific Expression of *MYB* intronic SNP rs210796

Comparing expression profiles of HPFH +/+ and HPFH -/- individuals has many limitations. We need to culture the cells from all individuals under the same exact conditions and use housekeeping genes as internal control to normalise for the differences that can be observed between cultures because of the influence from potential trans-acting factors. Comparing expression of the allele intra-individually overcomes many of these limitations by controlling the trans-acting environment and this enables detection of cis-acting regulatory haplotypes with far greater sensitivity. Ideally, allele specific expression is applied using exonic (coding) SNPs. We wanted to apply this approach to investigate allele specific expression of *MYB* but the absence of high frequency exonic SNPs in the MYB gene hindered this choice. Pastinen et al

suggested that intronic polymorphisms may be used to tag un-spliced heteronuclear RNA (hnRNA).

It has also been suggested that assays based on hnRNA provide a more specific test of transcriptional effects, as RNA-processing effects can be excluded but low abundance of hnRNA in total RNA preparations leads to “noise” in the assay. I screened four intronic SNPs (non-coding) of high frequency selected from the database (genome.ucsc.edu) to find appropriate SNPs for the allele specific expression experiments, and performed experiments using unspliced (primary) RNA (Figure 7.5).

Eventually the rs210796 (A/T) SNP in *MYB* intron 4 was selected based on its robustness in the assay. According to Hap Map data SNP rs210796 had a frequency of 0.456 increasing the likelihood of detecting heterozygous genotypes. A total of 50 healthy unrelated individuals of diverse ethnic backgrounds were recruited. DNA was isolated from peripheral blood of these individuals and genotyped for the relevant *HBS1L-MYB* intergenic variants, for the *MYB* intronic rs210796 SNP on chromosome 6q23 and for *HBS1L* exonic SNP rs13064 (as control) (See appendix for genotyping results).

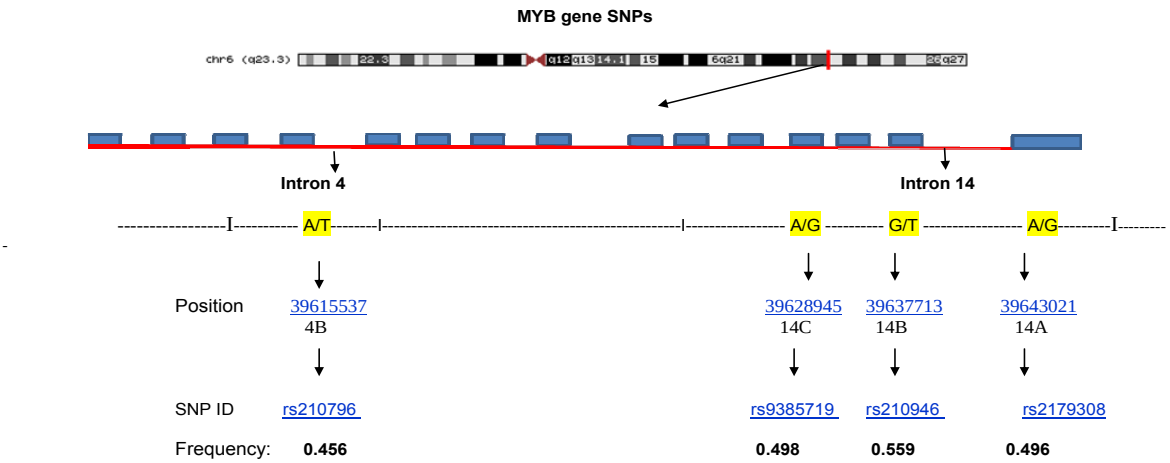


Figure 7.5: Informative SNP selection in *MYB* gene.
4 SNPs one from intron 4(rs210796) and three from intron 14 (rs9385719, rs210946 and rs2179308) were selected as informative SNPs. Each SNP had a frequency of > 0.45 increasing likelihood of heterozygosity.

7.2.3.1 Erythroid Cultures and Snapshot

Individuals with the appropriate intergenic and MYB intron 4 genotypes were selected for culture studies. SNP rs210796 (T/A) in *MYB* gene intron 4 was selected as the informative SNP to asses allele specific expression of *MYB* gene. Five healthy unrelated individuals heterozygous for both the *HMIP* block 2 haplotype i.e. (+/-) and *MYB* intronic SNP rs210796 were used as test samples. The (+) haplotype contains SNPs associated with elevated HbF; while the (-) haplotype does not contain any SNP associated with elevated HbF. Five healthy unrelated individuals homozygous for the HMIP block 2 haplotype(1 +/+ and 4 -/-) but heterozygous for *MYB* SNP rs210796 acted as controls for assessment of T:A expression imbalance.

RNA was isolated from phase II day 4 HEP cells and converted to cDNA with random hexamers instead of oligo dT since the SNP rs210796 was an intronic SNP. A fragment of 260bp encompassing rs210796 was specifically PCR amplified using genomic DNA and cDNA samples from the same individual. Genomic DNA samples were analysed by Snapshot in parallel with cDNA samples from the same individuals to assess relative expression of T and A alleles, and acted as controls. To control for variable amplification efficiency of T and A alleles, the T: A expression in cDNA samples were normalized in each individual internally with the genomic DNA T: A ratio and the T: A allele specific expression was derived. Experiments were performed in duplicates (Table 7.1 A).Five healthy unrelated individuals homozygous for the HMIP block 2 haplotype i.e. (+/+ or -/-) were used as controls for assessment of T:A expression imbalance. Individual 6 is *HMIP-2* +/+ while the other 4 controls are *HMIP-2* -/- (Table 7.1 B).The results suggest that individuals who are heterozygous for *HMIP*block two haplotype display greater imbalance in the relative expression of each allele when compared with individuals who are homozygous for the HMIP block 2 haplotype (Figure 7.6).

Test Samples (+/- genotype)	Technical Replicate 1			Technical Replicate 2			Average of normalized cDNA T:A
Sample ID	**DNA T/A**	**cDNA T/A**	**cDNA T/A**	**DNA T/A**	**cDNA T/A**	**cDNA T/A**	
			normalized			normalized	T:A
1	1.78	1.86	**1.04**	1.04	1.26	**1.21**	1.13
2	1.95	2.04	**1.05**	2.4	2.61	**1.09**	1.07
3	1.87	1.89	**1.01**	2.87	3.39	**1.19**	1.10
4	1.68	1.92	**1.14**	0.98	1.06	**1.08**	1.11
5	1.77	1.85	**1.05**	2.1	2.35	**1.12**	1.08

Controls (+/+ or -/- genotype)	Technical Replicate 1			Technical Replicate 2			Average of normalized cDNA T:A
Sample ID	**DNA T/A**	**cDNA T/A**	**cDNA T/A**	**DNA T/A**	**cDNA T/A**	**cDNA T/A**	
			normalized			normalized	T:A
6	1.68	1.52	**0.90**	2.13	2.07	**0.97**	0.94
7	1.8	1.67	**0.93**	2.25	2.17	**0.96**	0.94
8	1.88	1.93	**1.03**	2.17	1.87	**0.87**	0.95
9	1.81	1.69	**0.93**	2.78	2.2	**0.79**	0.86
10	1.78	1.61	**0.90**	2.39	2.34	**0.98**	0.94

Table7-1: Allele Specific Expression of *MYB* intronic SNP rs210796.
The T: A ratio of DNA samples and cDNA samples from both technical replicate is presented in the table. The T: A expression in cDNA samples were normalized in each individual internally with the genomic DNA T: A ratio.

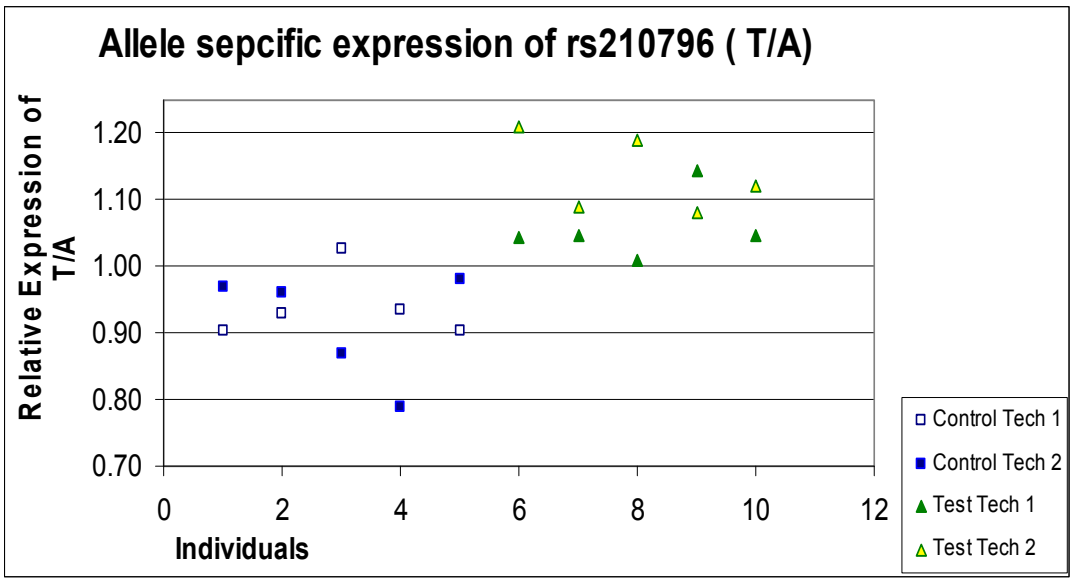


Figure7.6: Allele specific expression of *MYB* SNP rs210796 T/A.
Five healthy unrelated individuals whose genotype was +/- for the HMIP block 2 SNPs associated with elevated HbF were used as test samples. DNA and cDNA samples were analysed by Snapshot method for the relative expression of T and A alleles. The relative expression of T to A with cDNA samples were normalized to each individuals own T/A ratio with genomic DNA .Control samples refer to healthy unrelated individuals whose genotypes were +/+ or -/- for the HMIP block 2 SNPs associated with elevated HbF. Test samples refer to healthy unrelated individuals whose genotype was +/- for the HMIP block 2 SNPs associated with elevated HbF. Two sets of experiments were performed for each individual and both results are presented in the figure. T-test was performed (n: 10) and suggested that there is a significance difference between two groups (p: 0.03).

7.2.4 Allele Specific Expression of *HBS1L exonic* SNP rs13064

Allele specific expression of *HBS1L* was also investigated in relation to the *HMIP* block 2 variants. Presence of high frequency exonic SNP rs13064 (A/G) in *HBS1L* gene exon 17 enabled us to perform experiments using this coding SNP as our informative SNP. Individuals who were recruited for investigation of allele specific expression of MYB

were genotyped for SNP rs13064. The experiments should be performed with the same cDNA samples to limit inconsistencies coming from sample preparation.

The five individuals who were heterozygous for both the *HMIP* block 2 haplotype i.e. (+/-) and *MYB* intronic SNP rs210796, were genotyped for SNP rs13064 in *HBS1L*; but only 3 of the 5 individuals were heterozygous for SNP rs13064 (A/G) while other 2 individuals were homozygous for the major allele (A/A). The 5 healthy unrelated individuals who were homozygous for the *HMIP* block 2 haplotype i.e. (+/+ or -/-) were also genotyped and four individuals were heterozygous for *HBS1L* SNP rs13064 (A/G) while one individual was homozygous for the major allele (A/A).

The three healthy unrelated individuals heterozygous for the *HMIP* block 2 haplotype i.e. (+/-) were used as test samples. The (+) haplotype contains SNPs associated with elevated HbF; while the (-) haplotype does not contain any SNP associated with elevated HbF. All samples are heterozygous for this SNP. The cDNA samples were analysed by Snapshot in parallel with genomic DNA samples from same individuals to asses relative expression of A and G alleles. To control for variable amplification efficiency of A and G alleles, the A: G expression in cDNA samples were normalized in each individual internally with the genomic DNA A: G ratio and the A: G allele specific expression was derived (Table 7.2 A).The four healthy unrelated individuals homozygous for the HMIP block 2 haplotype i.e. (-/-) were used as controls for assessment of A:G expression imbalance. These individuals are also heterozygous for the *HBS1L* exonic rs13064 SNP (Table 7.2 B).

There was imbalanced expression of the A and G alleles both with controls and test samples, but there was no clear pattern with the A/G*HBS1L* allelic expression ratio ranging 0.40 to 1.60. The wide-range imbalanced expression was noted even in controls. The results suggest that the imbalanced expression may be related to other cis-acting polymorphisms that are not controlled in this experiment. Our sample size is limited because of restricted presence of heterozygotes for *HBS1L* exonic rs13064 SNP (Figure 7.7).

Test Samples (+/- genotype)			
Sample ID	**DNA A/G**	**cDNA A/G**	**cDNA A/G normalized**
6	1.31	0.58	**0.44**
13	1.19	0.97	**0.81**
14	0.93	0.6	**0.65**
Controls (-/- genotype)			
Sample ID	**DNA A/G**	**cDNA A/G**	**cDNA A/G normalized**
8	1.35	0.54	**0.40**
11	0.93	1.48	**1.58**
29	0.96	1.16	**1.21**
47	1.19	0.99	**0.84**

Table7-2: Allele Specific Expression of *HBS1L* exonic rs13064 SNP.
The A: G ratio of DNA samples and cDNA samples from single experiment is presented in the table. The A: G expressions in cDNA samples were normalized in each individual internally with the genomic DNA A: G ratio.

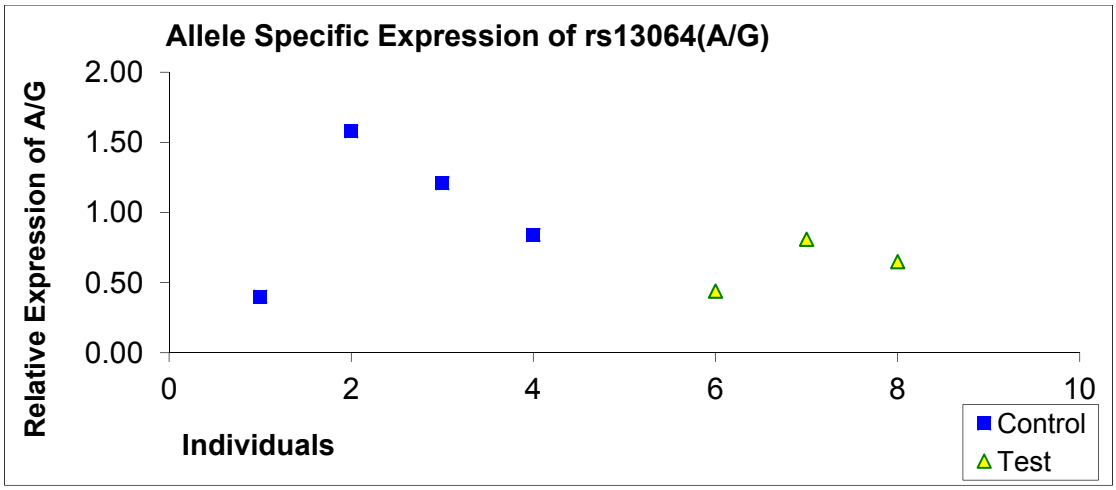


Figure7.7: Relative expression of A/G alleles of cDNA normalized to A/G ratio of genomic DNA is presented in the figure.
One set of experiment was performed from the erythroid RNA sample of each individual and the result is presented in the figure. Control samples refer to individuals whose genotypes were +/+ or -/- for the HMIP block 2 SNPs. Test samples refer to individuals whose genotype was +/- for the HMIP block 2 SNPs.

7.3 Discussion

A quantitative trait locus (QTL) controlling HbF levels was mapped to chromosome 6q23 in an Asian- Indian kindred with beta thalassaemia and heterocellular hereditary persistence of fetal hemoglobin (HPFH). The locus was refined to a region of 79 kbin the interval between *HBS1L* and *MYB* that has also been associated with other haematological traits.

Earlier on, before the interval was refined, Jiang et al demonstrated that *MYB* and *HBS1L* (two genes flanking the intergenic transcripts) are quantitative trait genes,with variable expression in erythroid progenitor cells of healthy adults. Analysis of a number of erythroid cultures of individuals has revealed a clear correlation of elevated HbF with low levels of *MYB* and *HBS1L* expression. It was hypothesized that the intergenic variants indirectly regulate HbF through influence of the expression of the flanking *MYB* and/or *HBS1L* genes by altering erythropoietic and cell cycle kinetics; but the mechanism of regulation remained elusive.

Recent genome-wide datasets generated by the ENCODE consortium were analysed to explore gene expression and intergenic regulatory potential within the *HBS1L-MYB* region for a variety of tissues. This analysis indicated that *MYB* expression was restricted to hematopoietic cells (erythroid K562 and lymphoid GM12878 cells) while *HBS1L* was expressed at similar levels in all cell types (ENCODE University of Washington DNase I Hypersensitivity by Digital DNase I; data not shown).

Further assessment of intergenic regulatory activity using a combination of genome-wide histone modification, DNaseI hypersensivity and genomic footprinting datasets suggested strong positive correlation between *MYB* expression levels and intergenic regulatory activity. In erythroid K562 cells, which express the highest levels of *MYB*, the

intergenic interval was transcribed and also contained numerous enhancer chromatin signatures and DNaseI hypersensitive sites (ENCODE University of Washington DNase I Hypersensitivity by Digital DNase I; data not shown).

Lymphoid Gm12878 cells expressing lower levels of *MYB* displayed fewer areas of regulatory activity and DNaseI hypersensitive sites. Cell types not expressing *MYB* (i.e. HUVEC, NHEK, and HepG2) displayed heterochromatinized or polycomb-repressed intergenic regions with an absence of DNaseI-hypersensitivity, while still expressing *HBS1L* at high levels.

Mukai et al reported a mouse model in which a transgene was inserted 77 kb upstream of the c-myb gene. The study was originally designed to examine the domain function of Erythropoietin receptor (EPOR) and the group generated three transgenic mouse lines bearing EPOR cDNA ligated to the *Gata1* gene-hematopoietic regulatory domain. In one of the transgenic mouse c-Myb expression was markedly decreased in megakaryocyte/erythrocyte lineage-restricted progenitors (MEPs) of the homozygous mutant mice leading to an imbalance between erythroid and megakaryocytic cellsdue to insertion mutagenesis and causing severe anaemia, thrombocythemia and splenomegaly. The group demonstrated that these haematological abnormalities did not correlate directly to the transgenic expression of EPOR since level of the EPOR transgene was about 64 times higher in another transgenic mouse which showed no haematological abnormalities. The group then suggested that the transgene insertion could block the function of an important enhancer for c-myb gene expression.The results clearly demonstrated that c-Myb is an essential regulator of the erythroid-megakaryocytic lineage bifurcation. Following on this, Suzuki et al confirmedthat insertion of this reported transgene in the intergenic region between the *Hbs1l* and *Myb* genes led to significant reduction of *Myb* expression, and reported slight reduction of *Hbs1l* expression.

The group also indicated that the mutant mouse exhibited typically elevated expression of embryonic globins and hematopoietic parameters similar to those observed in human HPFH and called this mouse a new mouse model of hereditary persistence of fetal haemoglobin (HPFH). Further expression profiling of key transcription factors have demonstrated direct correlation of *Hbs1l-Myb* locus disruption and TR2/TR4 (DRED pathway) expression and the group suggested that Myb activates DRED (direct repeat erythroid definitive) complex, a known repressor of embryonic and fetal globin genes in the adult. TR2/TR4 are repressors of embryonic globin genes. It was demonstrated that TR/TR4 recruit multiple epigenetic transcriptional corepressors such as DNA (cytosine-5)-methyltransferase 1 (DNMT1), nucleosome remodelling and histone deacetylase (NuRD), and lysine-specific demethylase-1 and its corepressor protein (LSD1/CoREST) and specially repress embryonic globin genes in differentiated erythroid progenitor cells (Cui et al 2011).

These results support the findings that variation in the *HBS1L-MYB* genomic domain is responsible for elevated expression of the fetal globin genes and suggest the involvement of Myb and DRED complex in this regulatory pathway.
The observations are in agreement with previous work by our group that demonstrated presence of regulatory elements in *HBS1L-MYB* intergenic region (*HMIP*) evidenced by GATA-1 binding coinciding with strong histone acetylation, RNA polymerase II activity, and erythroid specific DNase I hypersensitive sites. However, molecular insight into how this region regulates the flanking genes and which flanking gene is regulated remained elusive.

Studies indicated that variants associated with many diseases are localized in non-coding regulatory sequences that can regulate flanking genes though cis-acting influences on their expression (Bray et al 2003, Maurano et al 2012). This influence can be assessed by investigating the mRNA abundance. Assessing the relative expression levels of two SNP alleles (allelic imbalance) of a gene in the same sample,

instead of the total expression level as the quantitative phenotype is a powerful approach for identifying *cis*-acting regulatory SNPs or haplotypes. The major advantage of this approach is that the two SNP alleles are measured in the same environment, and serve as internal controls for each other. This excludes effects of environmental factors that may cause differences in the expression levels between samples.

A recent study by Schodel et al demonstrated how a polymorphism at a remote intergenic region on chromosome 11q13.3, a susceptibility locus for renal cell carcinoma, can lead to an allelic imbalance in cyclin D1 expression. This locus was identified by a GWAS study that investigatedpopulation-based cancer susceptibility loci; but until recently the mechanistic insights remained limited (Schodel et al 2012). The locus coincided with strong hypoxia-inducible factor (HIF) signals detected by ChIP-seq and the group identified two SNPs (rs7948643 and rs7939721) that lie only 10 bp and 15 bp from the HIF-binding motif at 11q13.3. The group showed that these variantsmodulates the binding and function of hypoxia-inducible factor (HIF), and also demonstrated that this region is a previously unrecognized transcriptional enhancer of *CCND1* (encoding cyclin D1) that is located 230 kb upstream of HIF binding site.

Further experiments demonstrated that modulation of HIF binding alter the allelic balance of *CCND1* expression. Using the allele specific expression approach the group identified two informative SNPs (rs7177 and rs678653) in the 3′ UTR of *CCND1* and demonstrated the allelic imbalance of *CCND1* expression. This study demonstrates the power of allele specific expression approach to provide insights into functions of variants in the intergenic regions (Schodel et al 2012).

We first constructed the expression profiles of each gene to select the appropriate stage for experiments. Healthy unrelated individuals were genotyped, and recruitments

with the appropriate genotypes were selected for culture studies. Our allelic expression experiments were performed using Snapshot-PCR, a technique that is quantitative.

Extensive optimization of the assays was carried out; and the Snapshot assays showed that alleles are not equally represented even with genomic DNA (where alleles are usually present in equimolar concentrations). This can be explained by a bias towards specific nucleotides (Moskvina et al 2005). Because of this, it was necessary to correct observed cDNA ratios by the ratios observed in genomic DNA, which, in the absence of copy number polymorphisms, should reflect a perfect 1: 1 ratio of the two alleles.

Our results suggested that individuals who are heterozygous for *HMIP-2* haplotype display greater imbalance in the relative expression of each *MYB* allele when compared with controls i.e. individuals who are homozygous for the HMIP block 2 haplotype (Figure 7.3). Departure from the defined 1: 1 ratio with cDNA samples (after normalization to genomic DNA) is suggested to reflect either epigenetic modification that results in preferential transcription from one parental chromosome or heterozygosity for DNA sequence variants that affect the transcription, splicing or stability of a gene. In the case of *HBS1L* the results were harder to interpret. Results suggested imbalanced expression of alleles both with controls and test samples. *HBS1L* allelic expression fluctuates from individual to individual from 0.40 to 1.60 even in controls. The results suggest that the imbalanced expression may be related to other cis-acting polymorphisms that are not controlled for in this experiment (Figure 7.4).

The direction of effect, in terms of allelic expression of the informative SNP, will depend upon the phase of the regulatory variant(s) with respect to the informative *MYB* SNP. In our case we have unrelated samples where phase between regulatory variant(s) with respect to the informative *MYB* SNP is not known.

Experiments should be repeated using erythroid progenitor material from the family members, who are heterozygotes for both HMIP block 2 SNPs and MYB intronic SNP rs210796 and in whom the linkage phase of these SNPs are known.

If the HMIP block 2 haplotype is associated with altered *HBS1L* expression, individuals who are heterozygous should display greater relative differences in the expression of each gene copy than those who are not. This relies on the assumption that no other regulatory haplotypes are present that influence expression of the gene.

It is possible that other potential variants present in control samples influence the *HBS1L* gene expression causing the observed imbalanced expression. The limitation of the experiment is the small sample size. Recruitment of appropriate individuals heterozygous for *HBS1L* exonic rs13064 SNP and *HMIP-2* was difficult.

MYB is a crucial factor with key roles in regulation of proliferation and cell lineage commitment of erythroid progenitor cells, thus its expression should be strictly regulated. The commitment of haematopoietic stem cells to the erythroid lineage is dependent on distinct threshold levels of MYB. We know from results that down regulation of *MYB* alter erythropoietic kinetics (Chapter 3); thus we were not expecting massive differences in the allele specific expression of *MYB* alleles. Association of the genetic variants in the *HMIP* variants with *MYB* control provides a functional link of the region with control of HbF and other haematological parameters.

Accelerated erythropoiesis as a response to relatively low MYB levels generates prematurely terminated erythroid cells that are still synthesising more HbF (Jiang et al 2006). We propose that the SNPs in the *HBS1L-MYB* intergenic region modulate *MYB* expression and this affects the variation in erythroid, platelet and monocyte counts as well as erythroid differentiation kinetics, the number of F cells and HbF.

Independently, Sankaran et al showed that downregulation of *MYB* in human erythroid progenitor cells resulted in increased expression of γ-globin and ε-globin. This study also demonstrated that *MYB* gene was a target for two microRNAs: miR-15a and miR-16-1, which are expressed in erythroid precursor cells. These miRNAs are overexpressed in trisomy 13 patients and the presence of high levels of miR-15a and miR-16-1 is the probable cause of the HbF levels, one of the features of the disease. *MYB* protein levels were reduced with even a slight over-expression of these miRNAs (Sankaran et al 2012). The influence of *MYB* on HbF levels were further supported by the findings that missense mutations in *MYB* strongly correlate with HbF (Galarneauet al 2010).

Altogether these studies (Jiang et al 2006, Galarneau et al 2010, and Sankaran et al 2012) demonstrate that MYB modulates HbF levels. We show that the intergenic variants modulate MYB levels providing the functional link for the genetic association of intergenic variants with HbF and F cells.

8 CONCLUSIONS AND FUTURE DIRECTIONS

The QTL on chromosome 6q23 was initially identified by linkage analysis in an extended Asian family with β-thalassaemia (Thein et al 1989). Subsequently, genetic variants in the *HBS1L-MYB* interval on chromosome 6q were shown to be highly associated, not only with HbF levels, but also linked to the control of other haematological parameters (Thein et al 2007, Menzel et al 2007). Genetic studies and resequencing of two key individuals from the family that led to the discovery of the 6q QTL revealed new SNPs in the *HBS1L-MYB* intergenic region that differ between individuals with HPFH and those without.This study was pursued to delineate key variants in the *HBS1L-MYB* intergenic region by performing functional experiments and the main conclusions were:

1. **Key transcription factor binding sites are present at the *HBS1L-MYB* intergenic region**

   In-silico predictions and previous ChIP-Chip experiments suggested presence of key transcription factor binding sites in the *HBS1L-MYB* intergenic region. ChIP-qPCR experiments confirmed and refined the GATA-1, KLF1 and MYB binding sites. Electrophoretic mobilty shift assays also confirmed the presence of these sites.

2. **High scoring variants located in two defined regions of *HMIP* block 2 that coincide with LDB1 binding sites alter binding of key transcription factors that may distrupt interactions with the *MYB* gene.**

   Electrophoretic Mobility Shift Assays (EMSAs) , Chromatin immunoprecipitation followed by real time PCR (ChIP-qPCR) and HaploChIP results suggested that variants altered the binding of key transcription factors such as GATA-1, KLF1 and MYB. The variants were located in two regions at 71kb and 84kb away from the *MYB* transcription start site and encompassed GATA-1 binding sites; coinciding with DNase I hypersensitive sites and intergenic transcripts.

Such enhancer regions were demonstrated in the homolgous region in mice. A study in mouse eryhtroid progenitors showed that these homologous regions physically interact with the promoter and first intron of the flanking *Myb* gene via chromatin looping suggesting that these two regions can be crucial in regulation of the flanking*Myb* gene (Stadhouders et al 2012).

3. **Erythroid culture kinetics of HPFH +/+ individuals is altered compared to that of HPFH -/- individuals.**

Comparisons of ex-vivo primary erythroid cell culture kinetics of HPFH +/+ individuals with HPFH -/- individuals showed that the kinetics of erythroid differentiation was noticeably different. HPFH +/+ erythroid progenitors differentiated much faster and monopoiesis was relativley higher in HPFH +/+ individuals compared to HPFH -/- individuals.

4. **There are strong indications that enhancer elements in the *HMIP- 2* regulate the flanking *MYB* but not *HBS1L gene*.**

Ex-vivo erythroid cell culture expression profiles in HPFH +/+ individuals compared to HPFH -/- individuals demonstrated that *MYB* expression was noticeably different; whereas expression *of HBS1L* ,the other flanking gene, was not. By comparing *MYB* expression intra-individually, we could confirm the imbalanced expression of *MYB* using the informative SNP rs201796 located in the intron 4 of *MYB* gene in HPFH +/- individuals.
Unfortunately, as the linkage phase of the SNPs in *HMIP-2* and the *MYB* intron 4 SNP rs201796 was not known, we could not determine the direction of the influence of the *HMIP-2* variants on *MYB* expression. Using the same approach in the same HPFH +/- individuals, we could not show imbalanced expression of *HBS1L.*

Taken together, our findings suggest the presence of a regulatory region in *HMIP-2* that regulates the flanking *MYB* gene distally. We have highlighted some functional variants that alter the binding of key transcription factors, providing further support to the importance of *MYB* in the regulation of erythroid development and demonstrated indications of a link between the intergenic variants and *MYB* expression. The results from this thesis were submitted to Journal of Clinical Investigation as a research article.

In 2006, Jiang et al demonstrated that *MYB* expression in *ex-vivo* erythroid progenitor cells of healthy adults is variable and individuals with high HbF had reduced MYB levels when compared to individuals with low HbF. Over expression of MYB in K562 cells resulted in low gamma globin gene expression. It was proposed then that MYB modulates HbF indirectly through the modulation of the number of F cells by altering the kinetics of erythroid differentiation. Accelerated erythropoiesis as a response to relatively low MYB levels could generate prematurely terminated erythroid cells synthesising more HbF (Jiang et al 2006).This proposal is not entirely speculative. It is already evident from previous studies of other groups that MYB is involved in cell proliferation and differentiation. The commitment of haematopoietic stem cells to the erythroid lineage is dependent on the distinct threshold levels of MYB. Reduced MYB levels favours commitment of haematopoietic progenitors to the megakaryocytic (platelet precursor) and monocytic lineages over the erythroid lineage (Emambokus et al 2003;Lu et al 2008). Our group also showed that *HMIP-2* genetic variants were associated with erythroid count and size, platelet and monocyte counts (Menzel et al 2007).

Thus, modulations in MYB levels could well underlie the variation in erythroid, platelet and monocyte counts as well as HbF levels associated with *HMIP*.

Later, genetic association studies by our group demonstrated the association of key variants in *HMIP-2* with HbF regulation (Thein et al 2007). Functional studies by Wahlberg et al indicated the presence of distal regulatory elements in *HBS1L-MYB*

intergenic region (*HMIP*) evidenced by GATA-1 binding coinciding with strong histone acetylation, RNA polymerase II activity, and erythroid specific DNase I hypersensitive sites (Wahlberg et al2009). It was proposed that the variants highly associated with HbF/ F cells in *HMIP-2* could regulate HbF indirectly via the control of its flanking *MYB* gene.

Several studies have related the 6q intergenic transcripts with altered *MYB* levels. The functional importance of the intergenic region was reported by Mukai et al who reported a mouse model in which a transgene bearing EPOR was inserted 77 kb upstream of the *myb* gene and caused insertion mutagenesis of the region.This led to marked decrease in *myb* expression in megakaryocyte/erythrocyte lineage-restricted progenitors (MEPs). Homozygous mutant mice were severely anaemic with a threefold increase in the platelet counts. The results suggested an association between the intergenic region and the *myb* gene; and clearly demonstrated that Myb is an essential regulator of the erythroid-megakaryocytic lineage bifurcation (Mukai et al 2006).

Analysis of recent ENCODE data also demonstrated a strong positive correlation between *MYB* expression levels and intergenic regulatory activity. In erythroid K562 cells, which express the highest levels of *MYB*, the intergenic interval was transcribed and contained numerous enhancer chromatin signatures and DNaseI hypersensitive sites.
On the other hand; cell types not expressing *MYB* for example, HUVEC, NHEK, and HepG2 displayed heterochromatinized or polycomb-repressed intergenic regions with an absence of DNaseI-hypersensitivity, while still expressing *HBS1L* at high levels. These results provide further support to the hypothesis that the region would preferentially be involved in *MYB* but not *HBS1L* regulation.

Results from this thesis clearly demonstrated a link between the 6q intergenic variants and *MYB* expression and provided additional insights into how these variants alter

MYB expression. We proposed that the region upstream of *MYB* contains distal regulatory elements occupied by key erythroid transcription factors that interact with *MYB* gene and form a key part of the overall control of *MYB* transcription. We showed that several of the variants alter transcription factor binding. The altered transcription factor binding can affect long range interactions with *MYB* gene by disrupting the complex including LDB1, GATA-1, and TAL1, ETO2 and KLF1 and leading to altered MYBexpression as discussed at Chapter 6 (Figure 8.1).

MYB has been associated with the control of HbF by various studies other than Jiang et al.Sankaran et al also showed that downregulation of *MYB* in erythroid progenitor cells resulted in increased expression of γ-globin and ε-globin (Sankaran et al 2012). Previous case studies suggested that raised HbF is a feature of trisomy 13 (Huehns et al1964). Sankaran *et al*(2011) mapped this elevation of HbF in Trisomy 13 cases to chromosomal band 13q14 which encodes two microRNAs: miR-15a and miR-16-1 expressed in erythroid precursor cells. Increasing the expression of these miRNAs in haematopoietic progenitor cells by 1.5 fold (as would be expected in a trisomy), resulted in an increase in γ-globin levels by 2.4 fold, which suggests that the elevated HbF levels seen in Trisomy 13 cases are likely due to the increased expression of these miRNAs. Interestingly, when the authors investigated the targets of these miRNAs, the *MYB* gene was found to have two conserved 8mer miR-15a/16-1 targeting sites. *MYB* protein levels were reduced with even a slight increase in the expression of these miRNAs suggesting that *MYB* gene influences the increased HbF levels observed in the trisomy 13 infants. The influence of *MYB* on HbF levels was further supported by the findings that missense mutations in *MYB* strongly correlate with HbF (Galarneauet al 2010).

Further studies of the mouse with the transgene insertion in the *Hbs1l-Myb* intergenic region showed that the position of the transgene coincided with the 24kb of *HMIP-2*. The transgene insertion led to a significant reduction of *Myb* gene expression, but only

to slight reduction of expression of the other flanking gene *HBS1L.* The mutant mouse exhibited typically elevated expression of embryonic globins and hematopoietic parameters similar to those observed in human HPFH. The group referred to the mutant mouse as a model of HPFH (Suzuki et al 2013).*HMIP-2* encompasses the two key regions -71kb and -84kb that physically interact with the promoter and first intron of the flanking *Myb* gene via chromatin looping suggesting that these two regions can be crucial in regulation of flanking genes (Stadhouders et al 2012).

Recently, other mechanisms whereby MYB can influence HbF levels have been proposed. A study in human erythroid progenitors suggests that MYB activates KLF1, in which case MYB can influence gamma globin expression via KLF1 activation of BCL11A (Bianchi et al 2010). Suzuki et al also demonstrated that reduction in *MYB* expression led to the downregulation of *KLF1* expression in mice which suggests a correlation between *MYB* and *KLF1* (Suzuki et al 2013).

Another mechanism of how MYB can influence HbF and gamma-globin gene expression is provided by the mouse model of HPFH in which the disruption of the*Hbs1l-Myb* locus led to a downregulation of TR2/TR4, nuclear receptors involved in the DRED complex.
The group suggested that Myb activates the DRED (direct repeat erythroid definitive) complex, a known repressor of embryonic and fetal globin genes in the adult (Suzuki et al 2013). These results support the findings that mutation of the *HBS1L-MYB* genomic domain is responsible for the elevated expression of the fetal globin genes via involvement of Myb and DRED complex in this regulatory pathway.

Overall, these results strongly indicate the functional importance of the intergenic region and suggest that the regions control HbF indirectly via the flanking *MYB* gene. Altered MYB expression can lead to the elevated HbF levels via a number of mechanisms:

- By altering erythropoietic kinetics that results in increased F cells and elevated HbF levels

- By altering expression of BCL11A, a known gamma globin repressor, via altering KLF1 expression

- By altering expression of TR2/TR4 (DRED pathway), another repressor of gamma globin expression (Figure 8.1).

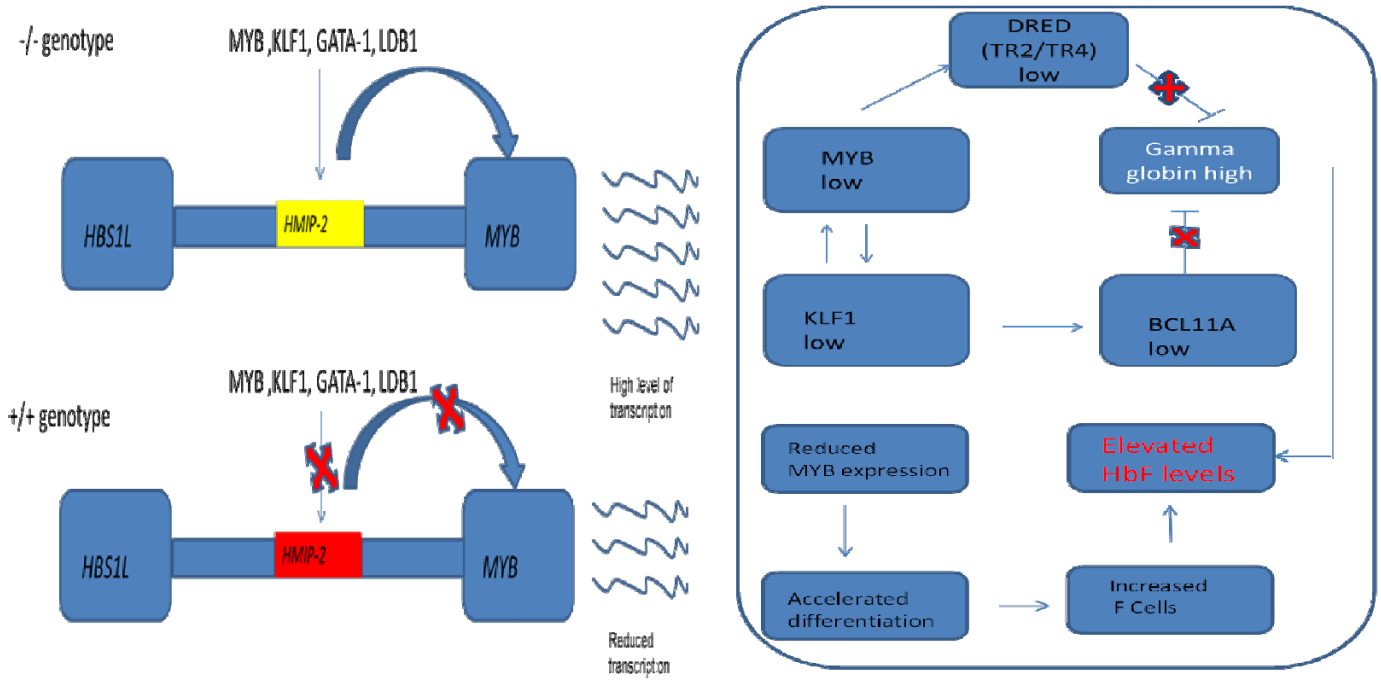


Figure 8.1: Potential mechanisms of how HMIP-2 variants alter HbF levels

8.1 Future Directions

8.1.1 Characterization of non-coding transcripts in the intergenic region

Previous tiling arrays revealed presence of transcription peaks at chromosome 6q QTL in erythroid progenitor cells (Wahlberg et al 2009). Transcripts were in two major peaks and were more recently confirmed in samples from other erythroid cultures. Previous studies by our group also showed that there are DNase 1 hypersensitive sites coinciding with these transcripts. Other studies also confirmed presence of DNase 1 hypersensitive sites at cd34+ progenitor cells around similar positions.

ENCODE data also showed that the intergenic interval is transcribed in erythroid K562 cells, which express the highest levels of *MYB*. The non-coding transcripts could be silenced to investigate their effects on *MYB* expression and to check the potential alterations at the kinetics of erythroid cultures. Such non-coding RNAs (ncRNA) have been known to contribute to gene regulation as well as physiological cellular processes.

Various non-coding RNAs are up-regulated at specific stages during the hematopoietic development and the functional relevance of non-coding RNAs has been proven at many different stages of lineage specification. Knockout of specific non-coding RNAs can produce dramatic phenotypes leading to severe hematopoietic defects. Several studies identified regions of intergenic transcription in the human β-globin locus. These intergenic transcripts coincided with DNase I hypersensitive sites and active histone modifications. It has been proposed that intergenic transcription in the globin gene cluster has a role in gene and chromatin activation (Haussecker et al 2005, Miles et al 2007).

Targeting the 6q intergenic transcripts with siRNAs could be a potential strategy to induce HbF via altering *MYB* expression. Another direct method of targeting MYB transcription factor with siRNA is possible but can have undesirable side effects, particularly given the pleiotropic role of *MYB* in haematopoiesis (Emambokus et al 2003; Carpinelli et al 2004). However, targeting intergenic region can lead to a partial knockdown of *MYB* and recent work suggests that such strategies may actually hold promise, because partial knockdown of *Myb* in mice had little effect on normal *in vivo* haematopoiesis while dramatically blocking progression of leukaemia (Zuber et al 2011). It is also possible to disrupt protein-DNA interactions by small molecules but it's a complicated task (Koehler et al 2010). The targeting of exogenous siRNAs to the desired cell type remains a major challenge, although on-going advances in siRNA delivery suggest that specific approaches to target erythroid cells may be developed in the future (Leuschner et al 2011).

8.1.2 Expression profiling of TR2/TR4

The recent study by Suzuki et al demonstrated that the intergenic region between *Hbs1l* and *Myb* is responsible for the persistence of murine embryonic globin gene in adult life.

They showed that the levels of expression of the flanking genes, *Hbs1l* and *Myb*, are reduced in the mutant erythroid progenitors. The erythroid progenitors of this transgenic mutant mouse also showeddiminished expression of the DNA binding subunits of the DRED complex (TR2/TR4).

These findings should be investigated in human erythroid progenitor cells. The correlation of *MYB* and TR2/TR4 expression should be investigated in erythroid progenitor cells of HPFH +/+ and HPFH -/- individuals that have low and high *MYB* expression, respectively.

8.1.3 Investigation of MYB transcription factor binding

Electrophoretic Mobility Shift Assays (EMSAs) and Chromatin immunoprecipitation followed by real time PCR (ChIP-qPCR) and HaploChIP showed the presence of MYB binding sites at the *HMIP-2* yet the experiments only focused on a limited area of the intergenic region coinciding with SNPs associated with HbF.

A broader study for MYB binding can be performed by MYB ChIP-chip with a custom array encompassing globin cluster, *KLF1, MYB,* and the intergenic region that would provide further insights in understanding its role in gene regulation and its association with HbF.

8.1.4 Characterization of transcription factor complex around rs66650371

EMSA experiments performed with oligonucleotide encompassing the three base pair deletion (rs66650371) demonstrated differential binding patterns with “wt” and “del” oligonucleotides (Chapter 5). To characterize protein-nucleic acid complexes, cold competition assays were performed but the results failed to suggest a relation between differentially bound complexes and GATA-1. There are other key transcription factor motifs that are encompassed by the oligonucleotides such as: RUNX1, Hepatic nuclear factor 1 (HNF1) (Chapter 4) and future experiments should be performed in order to characterize these protein-nucleic acid complexes. Further cold competition assays can be performed to investigate the presence of candidate transcription factors in the complex. If these experiments fail to suggest any correlation, experiments combining EMSA with mass spectroscopy could be performed to identify proteins in the complex.

8.1.5 Characterization of HBS1L-MYB intergenic region in other cell lineages

In this study, we have focused on the function of the *HMIP* block two in erythroid cells and only discussed its potential involvement in the control of *MYB* or *HBS1L* which are highly expressed in erythroid lineages. We should however not exclude the idea that the region could also influence expression of other genes in the locus in other cell lineages, especially monocytic lineage.

The mouse model of HPFH, with a transgenic insertion in the HBS1L-MYB intergenic region, also had reduced *Myb.* Interestingly, the same mouse model showed increased Myb mRNA in T-cells, suggesting that the transgenic insertion affected Myb expression differently in different haematopoietic lineages of the transgenic mouse (Mukai et al 2006).

Increased Myb levels have also been observed in feline T-cell lymphoma as a result of retroviral insertions in a region proximal to *HMIP* block 2 (Hanlon et al 2003). From these findings, it appears that the regulatory influence on *MYB* expression from this region differs between lineages. This suggests that the 6q intergenic variants may act as a *MYB* enhancer or silencer depending on the specific environment of the effector proteins (i.e. transcription factors and chromatin remodelling proteins) in the respective cell type. In the erythroid environment, the distal element appear to function as a control region involved in fine tuning of *MYB* levels required for normal haematopoietic lineage commitment and development.

9 REFERENCES

1. Alami, R., et al., *Deletions within the mouse beta-globin locus control region preferentially reduce beta(min) globin gene expression.* Genomics, 2000. **63**(3): p. 417-24.
2. Bank, A., *Understanding globin regulation in beta-thalassemia: it's as simple as alpha, beta, gamma, delta.* J Clin Invest, 2005. **115**(6): p. 1470-3.
3. Bank, A., *Regulation of human fetal hemoglobin: new players, new complexities.* Blood, 2006. **107**(2): p. 435-43.
4. Bartunek, P., et al., *GATA-1 and c-myb crosstalk during red blood cell differentiation through GATA-1 binding sites in the c-myb promoter.* Oncogene, 2003. **22**(13): p. 1927-35.
5. Bauer, A., et al., *The glucocorticoid receptor is required for stress erythropoiesis.* Genes Dev, 1999. **13**(22): p. 2996-3002.
6. Begley, C.G., et al., *The gene SCL is expressed during early hematopoiesis and encodes a differentiation-related DNA-binding motif.* Proc Natl Acad Sci U S A, 1989. **86**(24): p. 10128-32.
7. Bender, M.A., et al., *Beta-globin gene switching and DNase I sensitivity of the endogenous beta-globin locus in mice do not require the locus control region.* Mol Cell, 2000. **5**(2): p. 387-93.
8. Bender, T.P. and W.M. Kuehl, *Differential expression of the c-myb proto-oncogene marks the pre-B cell/B cell junction in murine B lymphoid tumors.* J Immunol, 1987. **139**(11): p. 3822-7.
9. Berger, S.L., *The complex language of chromatin regulation during transcription.* Nature, 2007. **447**(7143): p. 407-12.
10. Bhanu, N.V., et al., *A sustained and pancellular reversal of gamma-globin gene silencing in adult human erythroid precursor cells.* Blood, 2005. **105**(1): p. 387-93.
11. Bianchi, E., et al., *c-myb supports erythropoiesis through the transactivation of KLF1 and LMO2 expression.* Blood, 2010. **116**(22): p. e99-110.
12. Bieker, J.J., *Probing the onset and regulation of erythroid cell-specific gene expression.* Mt Sinai J Med, 2005. **72**(5): p. 333-8.
13. Borg, J., et al., *Haploinsufficiency for the erythroid transcription factor KLF1 causes hereditary persistence of fetal hemoglobin.* Nat Genet, 2010. **42**(9): p. 801-5.
14. Boyer, S.H., et al., *Fetal hemoglobin restriction to a few erythrocytes (F cells) in normal human adults.* Science, 1975. **188**(4186): p. 361-3.
15. Bray, N.J., et al., *Cis-acting variation in the expression of a high proportion of genes in human brain.* Hum Genet, 2003. **113**(2): p. 149-53.
16. Bray, N.J., et al., *Haplotypes at the dystrobrevin binding protein 1 (DTNBP1) gene locus mediate risk for schizophrenia through reduced DTNBP1 expression.* Hum Mol Genet, 2005. **14**(14): p. 1947-54.
17. Briegel, K., et al., *Regulation and function of transcription factor GATA-1 during red blood cell differentiation.* Development, 1996. **122**(12): p. 3839-50.
18. Broudy, V.C., *Stem cell factor and hematopoiesis.* Blood, 1997. **90**(4): p. 1345-64.
19. Bulger, M., et al., *Comparative structural and functional analysis of the olfactory receptor genes flanking the human and mouse beta-globin gene clusters.* Proc Natl Acad Sci U S A, 2000. **97**(26): p. 14560-5.
20. Campbell, A.D., et al., *Forced TR2/TR4 expression in sickle cell disease mice confers enhanced fetal hemoglobin synthesis and alleviated disease phenotypes.* Proc Natl Acad Sci U S A, 2011. **108**(46): p. 18808-13.
21. Cantor, A.B. and S.H. Orkin, *Coregulation of GATA factors by the Friend of GATA (FOG) family of multitype zinc finger proteins.* Semin Cell Dev Biol, 2005. **16**(1): p. 117-28.

22. Carpinelli, M.R., et al., *Suppressor screen in Mpl-/- mice: c-Myb mutation causes supraphysiological production of platelets in the absence of thrombopoietin signaling.* Proc Natl Acad Sci U S A, 2004. **101**(17): p. 6553-8.
23. Chasis, J.A. and N. Mohandas, *Erythroblastic islands: niches for erythropoiesis.* Blood, 2008. **112**(3): p. 470-8.
24. Cheng, Y., et al., *Erythroid GATA1 function revealed by genome-wide analysis of transcription factor occupancy, histone modifications, and mRNA expression.* Genome Res, 2009. **19**(12): p. 2172-84.
25. Clark, B.E. and S.L. Thein, *Molecular diagnosis of haemoglobin disorders.* Clin Lab Haematol, 2004. **26**(3): p. 159-76.
26. Close, J., et al., *Genome annotation of a 1.5 Mb region of human chromosome 6q23 encompassing a quantitative trait locus for fetal hemoglobin expression in adults.* BMC Genomics, 2004. **5**(1): p. 33.
27. Craig, J.E., et al., *Genetic heterogeneity in heterocellular hereditary persistence of fetal hemoglobin.* Blood, 1997. **90**(1): p. 428-34.
28. Creary, L.E., et al., *Genetic variation on chromosome 6 influences F cell levels in healthy individuals of African descent and HbF levels in sickle cell patients.* PLoS One, 2009. **4**(1): p. e4218.
29. Crossley, M., M. Merika, and S.H. Orkin, *Self-association of the erythroid transcription factor GATA-1 mediated by its zinc finger domains.* Mol Cell Biol, 1995. **15**(5): p. 2448-56.
30. Cui, S., et al., *Nuclear receptors TR2 and TR4 recruit multiple epigenetic transcriptional corepressors that associate specifically with the embryonic beta-type globin promoters in differentiated adult erythroid cells.* Mol Cell Biol, 2011. **31**(16): p. 3298-311.
31. Donze, D., T.M. Townes, and J.J. Bieker, *Role of erythroid Kruppel-like factor in human gamma- to beta-globin gene switching.* J Biol Chem, 1995. **270**(4): p. 1955-9.
32. Dooley, S., et al., *c-myb intron I protein binding and association with transcriptional activity in leukemic cells.* Leuk Res, 1996. **20**(5): p. 429-39.
33. Emambokus, N., et al., *Progression through key stages of haemopoiesis is dependent on distinct threshold levels of c-Myb.* EMBO J, 2003. **22**(17): p. 4478-88.
34. Enver, T., et al., *Developmental regulation of human fetal-to-adult globin gene switching in transgenic mice.* Nature, 1990. **344**(6264): p. 309-13.
35. Evans, T. and G. Felsenfeld, *The erythroid-specific transcription factor Eryf1: a new finger protein.* Cell, 1989. **58**(5): p. 877-85.
36. Farrell, J. J., R. M. Sherva, et al. (2011). "A 3-bp deletion in the HBS1L-MYB intergenic region on chromosome 6q23 is associated with HbF expression." Blood 117(18): 4935-4945.
37. Fell, H.P., R.G. Smith, and P.W. Tucker, *Molecular analysis of the t(2;14) translocation of childhood chronic lymphocytic leukemia.* Science, 1986. **232**(4749): p. 491-4.
38. Feng, W.C., C.M. Southwood, and J.J. Bieker, *Analyses of beta-thalassemia mutant DNA interactions with erythroid Kruppel-like factor (EKLF), an erythroid cell-specific transcription factor.* J Biol Chem, 1994. **269**(2): p. 1493-500.
39. Fibach, E., et al., *Hydroxyurea increases fetal hemoglobin in cultured erythroid cells derived from normal individuals and patients with sickle cell anemia or beta-thalassemia.* Blood, 1993. **81**(6): p. 1630-5.
40. Fibach, E., et al., *Proliferation and maturation of human erythroid progenitors in liquid culture.* Blood, 1989. **73**(1): p. 100-3.
41. Fire, A., et al., *Potent and specific genetic interference by double-stranded RNA in Caenorhabditis elegans.* Nature, 1998. **391**(6669): p. 806-11.
42. Forrester, W.C., et al., *Evidence for a locus activation region: the formation of developmentally stable hypersensitive sites in globin-expressing hybrids.* Nucleic Acids Res, 1987. **15**(24): p. 10159-77.

43. Forrester, W.C., et al., *A developmentally stable chromatin structure in the human beta-globin gene cluster.* Proc Natl Acad Sci U S A, 1986. **83**(5): p. 1359-63.
44. Friedman, A.D., *Runx1, c-Myb, and C/EBPalpha couple differentiation to proliferation or growth arrest during hematopoiesis.* J Cell Biochem, 2002. **86**(4): p. 624-9.
45. Fritsch, E.F., R.M. Lawn, and T. Maniatis, *Molecular cloning and characterization of the human beta-like globin gene cluster.* Cell, 1980. **19**(4): p. 959-72.
46. Galarneau, G., et al., *Fine-mapping at three loci known to affect fetal hemoglobin levels explains additional genetic variation.* Nat Genet, 2010. **42**(12): p. 1049-51.
47. Galle, C., et al., *Predominance of type 1 CD4+ T cells in human abdominal aortic aneurysm.* Clin Exp Immunol, 2005. **142**(3): p. 519-27.
48. Garner, C., et al., *Heterocellular hereditary persistence of fetal haemoglobin affects the haematological parameters of beta-thalassaemia trait.* Br J Haematol, 2003. **123**(2): p. 353-8.
49. Garner, C., et al., *Quantitative trait locus on chromosome 8q influences the switch from fetal to adult hemoglobin.* Blood, 2004. **104**(7): p. 2184-6.
50. Garner, C., et al., *Genetic influences on F cells and other hematologic variables: a twin heritability study.* Blood, 2000. **95**(1): p. 342-6.
51. Garrick, D., et al., *Switching genes on and off in haemopoiesis.* Biochem Soc Trans, 2008. **36**(Pt 4): p. 613-8.
52. Gonda, T.J., D.K. Sheiness, and J.M. Bishop, *Transcripts from the cellular homologs of retroviral oncogenes: distribution among chicken tissues.* Mol Cell Biol, 1982. **2**(6): p. 617-24.
53. Gong, Q.H., J.C. McDowell, and A. Dean, *Essential role of NF-E2 in remodeling of chromatin structure and transcriptional activation of the epsilon-globin gene in vivo by 5' hypersensitive site 2 of the beta-globin locus control region.* Mol Cell Biol, 1996. **16**(11): p. 6055-64.
54. Goodwin, A.J., et al., *In vivo formation of a human beta-globin locus control region core element requires binding sites for multiple factors including GATA-1, NF-E2, erythroid Kruppel-like factor, and Sp1.* J Biol Chem, 2001. **276**(29): p. 26883-92.
55. Gregory, C.J. and A.C. Eaves, *Three stages of erythropoietic progenitor cell differentiation distinguished by a number of physical and biologic properties.* Blood, 1978. **51**(3): p. 527-37.
56. Greig, K.T., S. Carotta, and S.L. Nutt, *Critical roles for c-Myb in hematopoietic progenitor cells.* Semin Immunol, 2008. **20**(4): p. 247-56.
57. Grosveld, F., et al., *The regulation of human globin gene switching.* Philos Trans R Soc Lond B Biol Sci, 1993. **339**(1288): p. 183-91.
58. Grosveld, F., et al., *The regulation of expression of human beta-globin genes.* Prog Clin Biol Res, 1987. **251**: p. 133-44.
59. Halupa, A., et al., *Erythropoietin receptor Y479 couples to ERK1/2 activation via recruitment of phospholipase Cgamma.* Exp Cell Res, 2005. **309**(1): p. 1-11.
60. Hanlon, L., et al., *Long-range effects of retroviral insertion on c-myb: overexpression may be obscured by silencing during tumor growth in vitro.* J Virol, 2003. **77**(2): p. 1059-68.
61. Harju, S., K.J. McQueen, and K.R. Peterson, *Chromatin structure and control of beta-like globin gene switching.* Exp Biol Med (Maywood), 2002. **227**(9): p. 683-700.
62. Haussecker, D. and N.J. Proudfoot, *Dicer-dependent turnover of intergenic transcripts from the human beta-globin gene cluster.* Mol Cell Biol, 2005. **25**(21): p. 9724-33.
63. Hellman, L.M. and M.G. Fried, *Electrophoretic mobility shift assay (EMSA) for detecting protein-nucleic acid interactions.* Nat Protoc, 2007. **2**(8): p. 1849-61.
64. Higgs, D.R., *The thalassaemia syndromes.* Q J Med, 1993. **86**(9): p. 559-64.

65. Higgs, D.R., *alpha-Thalassaemia.* Baillieres Clin Haematol, 1993. **6**(1): p. 117-50.
66. Higgs, D.R., *Gene regulation in hematopoiesis: new lessons from thalassemia.* Hematology Am Soc Hematol Educ Program, 2004: p. 1-13.
67. Higgs, D.R. and R.J. Gibbons, *The molecular basis of alpha-thalassemia: a model for understanding human molecular genetics.* Hematol Oncol Clin North Am. **24**(6): p. 1033-54.
68. Higgs, D.R., D. Vernimmen, and B. Wood, *Long-range regulation of alpha-globin gene expression.* Adv Genet, 2008. **61**: p. 143-73.
69. Higgs, D.R. and D.J. Weatherall, *The alpha thalassaemias.* Cell Mol Life Sci, 2009. **66**(7): p. 1154-62.
70. Higgs, D.R., et al., *A major positive regulatory region located far upstream of the human alpha-globin gene locus.* Genes Dev, 1990. **4**(9): p. 1588-601.
71. Ho, P.J. and S.L. Thein, *Gene regulation and deregulation: a beta globin perspective.* Blood Rev, 2000. **14**(2): p. 78-93.
72. Hong, W., et al., *FOG-1 recruits the NuRD repressor complex to mediate transcriptional repression by GATA-1.* EMBO J, 2005. **24**(13): p. 2367-78.
73. Huehns, E.R., M. Lutzner, and F. Hecht, *Nuclear Abnormalities of the Neutrophils in D1 (13-15)-Trisomy Syndrome.* Lancet, 1964. **1**(7333): p. 589-90.
74. Jawaid, K., et al., *Binding patterns of BCL11A in the globin and GATA1 loci and characterization of the BCL11A fetal hemoglobin locus.* Blood Cells Mol Dis, 2010. **45**(2): p. 140-6.
75. Jiang, J., et al., *cMYB is involved in the regulation of fetal hemoglobin production in adults.* Blood, 2006. **108**(3): p. 1077-83.
76. Kollias, G., et al., *The human beta-globin gene contains a downstream developmental specific enhancer.* Nucleic Acids Res, 1987. **15**(14): p. 5739-47.
77. Layon, M.E., et al., *Expression of GATA-1 in a non-hematopoietic cell line induces beta-globin locus control region chromatin structure remodeling and an erythroid pattern of gene expression.* J Mol Biol, 2007. **366**(3): p. 737-44.
78. Leberbauer, C., et al., *Different steroids co-regulate long-term expansion versus terminal differentiation in primary human erythroid progenitors.* Blood, 2005. **105**(1): p. 85-94.
79. Leimberg, J.M., A.M. Konijn, and E. Fibach, *Macrophages promote development of human erythroid precursors in transferrin-free culture medium.* Hematology, 2005. **10**(1): p. 73-6.
80. Lettre, G., et al., *DNA polymorphisms at the BCL11A, HBS1L-MYB, and beta-globin loci associate with fetal hemoglobin levels and pain crises in sickle cell disease.* Proc Natl Acad Sci U S A, 2008. **105**(33): p. 11869-74.
81. Leuschner, F., et al., *Therapeutic siRNA silencing in inflammatory monocytes in mice.* Nat Biotechnol, 2011. **29**(11): p. 1005-10.
82. Ley, T.J., et al., *5-azacytidine selectively increases gamma-globin synthesis in a patient with beta+ thalassemia.* N Engl J Med, 1982. **307**(24): p. 1469-75.
83. Lin, H.H., et al., *Functional analysis of the c-myb proto-oncogene.* Curr Top Microbiol Immunol, 1996. **211**: p. 79-87.
84. Lipsick, J.S., *One billion years of Myb.* Oncogene, 1996. **13**(2): p. 223-35.
85. Liu, H., et al., *Functional studies of BCL11A: characterization of the conserved BCL11A-XL splice variant and its interaction with BCL6 in nuclear paraspeckles of germinal center B cells.* Mol Cancer, 2006. **5**: p. 18.
86. Liu, P., et al., *Bcl11a is essential for normal lymphoid development.* Nat Immunol, 2003. **4**(6): p. 525-32.
87. Lu, J., et al., *MicroRNA-mediated control of cell fate in megakaryocyte-erythrocyte progenitors.* Dev Cell, 2008. **14**(6): p. 843-53.
88. Mahajan, M.C., S. Karmakar, and S.M. Weissman, *Control of beta globin genes.* J Cell Biochem, 2007. **102**(4): p. 801-10.
89. Makani, J., et al., *Genetics of fetal hemoglobin in Tanzanian and British patients with sickle cell anemia.* Blood, 2011. **117**(4): p. 1390-2.

90. Maniatis, T., et al., *The molecular genetics of human hemoglobins.* Annu Rev Genet, 1980. **14**: p. 145-78.
91. Maurano, M.T., et al., *Systematic localization of common disease-associated variation in regulatory DNA.* Science, 2012. **337**(6099): p. 1190-5.
92. McDonald, W.H., et al., *Myb-related fission yeast cdc5p is a component of a 40S snRNP-containing complex and is essential for pre-mRNA splicing.* Mol Cell Biol, 1999. **19**(8): p. 5352-62.
93. Menzel, S., et al., *A QTL influencing F cell production maps to a gene encoding a zinc-finger protein on chromosome 2p15.* Nat Genet, 2007. **39**(10): p. 1197-9.
94. Menzel, S., et al., *The HBS1L-MYB intergenic region on chromosome 6q23.3 influences erythrocyte, platelet, and monocyte counts in humans.* Blood, 2007. **110**(10): p. 3624-6.
95. Menzel, S., et al., *Experimental generation of SNP haplotype signatures in patients with sickle cell anaemia.* PLoS One. **5**(9): p. e13004.
96. Merika, M. and S.H. Orkin, *Functional synergy and physical interactions of the erythroid transcription factor GATA-1 with the Kruppel family proteins Sp1 and EKLF.* Mol Cell Biol, 1995. **15**(5): p. 2437-47.
97. Mikkola, H.K. and S.H. Orkin, *The journey of developing hematopoietic stem cells.* Development, 2006. **133**(19): p. 3733-44.
98. Miles, J., et al., *Intergenic transcription, cell-cycle and the developmentally regulated epigenetic profile of the human beta-globin locus.* PLoS One, 2007. **2**(7): p. e630.
99. Miller, I.J. and J.J. Bieker, *A novel, erythroid cell-specific murine transcription factor that binds to the CACCC element and is related to the Kruppel family of nuclear proteins.* Mol Cell Biol, 1993. **13**(5): p. 2776-86.
100. Miller, J.L., *Signaled expression of fetal hemoglobin during development.* Transfusion, 2005. **45**(7): p. 1229-32.
101. Milot, E., et al., *Heterochromatin effects on the frequency and duration of LCR-mediated gene transcription.* Cell, 1996. **87**(1): p. 105-14.
102. Morceau, F., et al., *GATA-1: friends, brothers, and coworkers.* Ann N Y Acad Sci, 2004. **1030**: p. 537-54.
103. Moreau-Gaudry, F., et al., *High-level erythroid-specific gene expression in primary human and murine hematopoietic cells with self-inactivating lentiviral vectors.* Blood, 2001. **98**(9): p. 2664-72.
104. Moskvina, V., et al., *Design of case-controls studies with unscreened controls.* Ann Hum Genet, 2005. **69**(Pt 5): p. 566-76.
105. Mucenski, M.L., et al., *A functional c-myb gene is required for normal murine fetal hepatic hematopoiesis.* Cell, 1991. **65**(4): p. 677-89.
106. Mukai, H.Y., et al., *Transgene insertion in proximity to the c-myb gene disrupts erythroid-megakaryocytic lineage bifurcation.* Mol Cell Biol, 2006. **26**(21): p. 7953-65.
107. Ney, P.A., et al., *Inducibility of the HS II enhancer depends on binding of an erythroid specific nuclear protein.* Nucleic Acids Res, 1990. **18**(20): p. 6011-7.
108. Nicolaides, N.C., et al., *Positive autoregulation of c-myb expression via Myb binding sites in the 5' flanking region of the human c-myb gene.* Mol Cell Biol, 1991. **11**(12): p. 6166-76.
109. Noguchi, C.T., A.N. Schechter, and G.P. Rodgers, *Sickle cell disease pathophysiology.* Baillieres Clin Haematol, 1993. **6**(1): p. 57-91.
110. Noordermeer, D., et al., *Transcription and chromatin organization of a housekeeping gene cluster containing an integrated beta-globin locus control region.* PLoS Genet, 2008. **4**(3): p. e1000016.
111. Noordermeer, D. and W. de Laat, *Joining the loops: beta-globin gene regulation.* IUBMB Life, 2008. **60**(12): p. 824-33.
112. Nuez, B., et al., *Defective haematopoiesis in fetal liver resulting from inactivation of the EKLF gene.* Nature, 1995. **375**(6529): p. 316-8.

113. Nuinoon, M., et al., *A genome-wide association identified the common genetic variants influence disease severity in beta0-thalassemia/hemoglobin E.* Hum Genet. **127**(3): p. 303-14.
114. Ogawa, M., et al., *Expression and function of c-Kit in fetal hemopoietic progenitor cells: transition from the early c-Kit-independent to the late c-Kit-dependent wave of hemopoiesis in the murine embryo.* Development, 1993. **117**(3): p. 1089-98.
115. Orkin, S.H., *GATA-binding transcription factors in hematopoietic cells.* Blood, 1992. **80**(3): p. 575-81.
116. Ottersbach, K. and E. Dzierzak, *The murine placenta contains hematopoietic stem cells within the vascular labyrinth region.* Dev Cell, 2005. **8**(3): p. 377-87.
117. Pandit, R.A., et al., *Association of SNP in exon 1 of HBS1L with hemoglobin F level in beta0-thalassemia/hemoglobin E.* Int J Hematol, 2008. **88**(4): p. 357-61.
118. Papayannopoulou, T., et al., *Globin gene switching: insights from studies in somatic heterospecific hybrids and in transgenic mice.* Prog Clin Biol Res, 1990. **352**: p. 323-8.
119. Papayannopoulou, T., et al., *Fetal to adult hemopoietic cell transplantation in humans: insights into hemoglobin switching.* Blood, 1986. **67**(1): p. 99-104.
120. Pauling, L., H.A. Itano, and et al., *Sickle cell anemia a molecular disease.* Science, 1949. **110**(2865): p. 543-8.
121. Pecot, C.V., et al., *RNA interference in the clinic: challenges and future directions.* Nat Rev Cancer, 2011. **11**(1): p. 59-67.
122. Perkins, A.C., A.H. Sharpe, and S.H. Orkin, *Lethal beta-thalassaemia in mice lacking the erythroid CACCC-transcription factor EKLF.* Nature, 1995. **375**(6529): p. 318-22.
123. Perutz, M.F., *Structure of hemoglobin.* Brookhaven Symp Biol, 1960. **13**: p. 165-83.
124. Pomerantz, O., et al., *Conserved elements containing NF-E2 and tandem GATA binding sites are required for erythroid-specific chromatin structure reorganization within the human beta-globin locus control region.* Nucleic Acids Res, 1998. **26**(24): p. 5684-91.
125. Porcher, C., et al., *The T cell leukemia oncoprotein SCL/tal-1 is essential for development of all hematopoietic lineages.* Cell, 1996. **86**(1): p. 47-57.
126. Quandt, K., K. Frech, and T. Werner, *[Analysis of transcriptional regulatory regions based on detecting transcription factor binding sites and studies of their mutual position].* Mol Biol (Mosk), 1997. **31**(4): p. 749-58.
127. Ragoczy, T., et al., *The locus control region is required for association of the murine beta-globin locus with engaged transcription factories during erythroid maturation.* Genes Dev, 2006. **20**(11): p. 1447-57.
128. Revzin, A., *Gel electrophoresis assays for DNA-protein interactions.* Biotechniques, 1989. **7**(4): p. 346-55.
129. Rochette, J., et al., *Association of a novel high oxygen affinity haemoglobin variant with delta beta thalassaemia.* Br J Haematol, 1994. **86**(1): p. 118-24.
130. Rochette, J., J.E. Craig, and S.L. Thein, *Fetal hemoglobin levels in adults.* Blood Rev, 1994. **8**(4): p. 213-24.
131. Sankaran, V.G., et al., *MicroRNA-15a and -16-1 act via MYB to elevate fetal hemoglobin expression in human trisomy 13.* Proc Natl Acad Sci U S A, 2011. **108**(4): p. 1519-24.
132. Sankaran, V.G., et al., *Human fetal hemoglobin expression is regulated by the developmental stage-specific repressor BCL11A.* Science, 2008. **322**(5909): p. 1839-42.
133. Sankaran, V.G., J. Xu, and S.H. Orkin, *Transcriptional silencing of fetal hemoglobin by BCL11A.* Ann N Y Acad Sci, 2010. **1202**: p. 64-8.
134. Sankaran, V.G., et al., *Developmental and species-divergent globin switching are driven by BCL11A.* Nature, 2009. **460**(7259): p. 1093-7.
135. Sathyanarayana, P., et al., *EPO receptor circuits for primary erythroblast survival.* Blood, 2008. **111**(11): p. 5390-9.

136. Schodel, J., et al., *Common genetic variants at the 11q13.3 renal cancer susceptibility locus influence binding of HIF to an enhancer of cyclin D1 expression.* Nat Genet, 2012. **44**(4): p. 420-5, S1-2.
137. Schrier, S.L., *Thalassemia: pathophysiology of red cell changes.* Annu Rev Med, 1994. **45**: p. 211-8.
138. Schrier, S.L., *Pathophysiology of thalassemia.* Curr Opin Hematol, 2002. **9**(2): p. 123-6.
139. Schrier, S.L. and N. Mohandas, *Globin-chain specificity of oxidation-induced changes in red blood cell membrane properties.* Blood, 1992. **79**(6): p. 1586-92.
140. Schrier, S.L., E. Rachmilewitz, and N. Mohandas, *Cellular and membrane properties of alpha and beta thalassemic erythrocytes are different: implication for differences in clinical manifestations.* Blood, 1989. **74**(6): p. 2194-202.
141. Seib, T., et al., *Presence of regulatory sequences within intron 4 of human and murine c-myb genes.* Biochim Biophys Acta, 1994. **1219**(2): p. 285-92.
142. So, C.C., et al., *The HBS1L-MYB intergenic region on chromosome 6q23 is a quantitative trait locus controlling fetal haemoglobin level in carriers of beta-thalassaemia.* J Med Genet, 2008. **45**(11): p. 745-51.
143. Stadhouders, R., et al., *Dynamic long-range chromatin interactions control Myb proto-oncogene transcription during erythroid development.* EMBO J, 2012. **31**(4): p. 986-99.
144. Stamatoyannopoulos, G., *The molecular basis of hemoglobin disease.* Annu Rev Genet, 1972. **6**: p. 47-70.
145. Stamatoyannopoulos, G., *Prospects for developing a molecular cure for thalassemia.* Hematology, 2005. **10 Suppl 1**: p. 255-7.
146. Stamatoyannopoulos, G., *Control of globin gene expression during development and erythroid differentiation.* Exp Hematol, 2005. **33**(3): p. 259-71.
147. Stuart, M.J. and R.L. Nagel, *Sickle-cell disease.* Lancet, 2004. **364**(9442): p. 1343-60.
148. Sullivan, J., et al., *Identification of the major positive regulators of c-myb expression in hematopoietic cells of different lineages.* J Biol Chem, 1997. **272**(3): p. 1943-9.
149. Sumner, R., et al., *Initiation of adult myelopoiesis can occur in the absence of c-Myb whereas subsequent development is strictly dependent on the transcription factor.* Oncogene, 2000. **19**(30): p. 3335-42.
150. Suzuki, M., et al., *Disruption of the Hbs1l-Myb locus causes hereditary persistence of fetal hemoglobin in a mouse model.* Mol Cell Biol, 2013. **33**(8): p. 1687-95.
151. Talbot, D. and F. Grosveld, *The 5'HS2 of the globin locus control region enhances transcription through the interaction of a multimeric complex binding at two functionally distinct NF-E2 binding sites.* EMBO J, 1991. **10**(6): p. 1391-8.
152. Tallack, M.R. and A.C. Perkins, *KLF1 directly coordinates almost all aspects of terminal erythroid differentiation.* IUBMB Life, 2010. **62**(12): p. 886-90.
153. Tewari, R., et al., *The human beta-globin locus control region confers an early embryonic erythroid-specific expression pattern to a basic promoter driving the bacterial lacZ gene.* Development, 1996. **122**(12): p. 3991-9.
154. Tewari, R., et al., *Erythroid Kruppel-like factor (EKLF) is active in primitive and definitive erythroid cells and is required for the function of 5'HS3 of the beta-globin locus control region.* EMBO J, 1998. **17**(8): p. 2334-41.
155. Thein, S.L., *Beta-thalassaemia.* Baillieres Clin Haematol, 1998. **11**(1): p. 91-126.
156. Thein, S.L., *Genetic modifiers of the beta-haemoglobinopathies.* Br J Haematol, 2008. **141**(3): p. 357-66.
157. Thein, S.L. and J.E. Craig, *Genetics of Hb F/F cell variance in adults and heterocellular hereditary persistence of fetal hemoglobin.* Hemoglobin, 1998. **22**(5-6): p. 401-14.
158. Thein, S.L. and S. Menzel, *Discovering the genetics underlying foetal haemoglobin production in adults.* Br J Haematol, 2009. **145**(4): p. 455-67.

159. Thein, S.L., et al., *Control of fetal hemoglobin: new insights emerging from genomics and clinical implications.* Hum Mol Genet, 2009. **18**(R2): p. R216-23.
160. Thein, S.L., et al., *Intergenic variants of HBS1L-MYB are responsible for a major quantitative trait locus on chromosome 6q23 influencing fetal hemoglobin levels in adults.* Proc Natl Acad Sci U S A, 2007. **104**(27): p. 11346-51.
161. Thein, S.L., et al., *Detection of a major gene for heterocellular hereditary persistence of fetal hemoglobin after accounting for genetic modifiers.* Am J Hum Genet, 1994. **54**(2): p. 214-28.
162. Thein, S.L., et al., *Association of thalassaemia intermedia with a beta-globin gene haplotype.* Br J Haematol, 1987. **65**(3): p. 367-73.
163. Thein, S.L. and D.J. Weatherall, *A non-deletion hereditary persistence of fetal hemoglobin (HPFH) determinant not linked to the beta-globin gene complex.* Prog Clin Biol Res, 1989. **316B**: p. 97-111.
164. Tolhuis, B., et al., *Looping and interaction between hypersensitive sites in the active beta-globin locus.* Mol Cell, 2002. **10**(6): p. 1453-65.
165. Tsai, F.Y., et al., *An early haematopoietic defect in mice lacking the transcription factor GATA-2.* Nature, 1994. **371**(6494): p. 221-6.
166. Tuan, D., et al., *The "beta-like-globin" gene domain in human erythroid cells.* Proc Natl Acad Sci U S A, 1985. **82**(19): p. 6384-8.
167. Uda, M., et al., *Genome-wide association study shows BCL11A associated with persistent fetal hemoglobin and amelioration of the phenotype of beta-thalassemia.* Proc Natl Acad Sci U S A, 2008. **105**(5): p. 1620-5.
168. Vakoc, C.R., et al., *Proximity among distant regulatory elements at the beta-globin locus requires GATA-1 and FOG-1.* Mol Cell, 2005. **17**(3): p. 453-62.
169. van den Akker, E., et al., *The majority of the in vitro erythroid expansion potential resides in CD34(-) cells, outweighing the contribution of CD34(+) cells and significantly increasing the erythroblast yield from peripheral blood samples.* Haematologica, 2010. **95**(9): p. 1594-8.
170. Vieira, K.F., et al., *Recruitment of transcription complexes to the beta-globin gene locus in vivo and in vitro.* J Biol Chem, 2004. **279**(48): p. 50350-7.
171. Wadman, I.A., et al., *The LIM-only protein Lmo2 is a bridging molecule assembling an erythroid, DNA-binding complex which includes the TAL1, E47, GATA-1 and Ldb1/NLI proteins.* Embo J, 1997. **16**(11): p. 3145-57.
172. Wahlberg, K., et al., *The HBS1L-MYB intergenic interval associated with elevated HbF levels shows characteristics of a distal regulatory region in erythroid cells.* Blood, 2009. **114**(6): p. 1254-62.
173. Wallrapp, C., et al., *The product of the mammalian orthologue of the Saccharomyces cerevisiae HBS1 gene is phylogenetically related to eukaryotic release factor 3 (eRF3) but does not carry eRF3-like activity.* FEBS Lett, 1998. **440**(3): p. 387-92.
174. Weatherall, D.J., *The phenotypic diversity of monogenic disease: lessons from the thalassemias.* Harvey Lect, 1998. **94**: p. 1-20.
175. Weatherall, D.J., *Phenotype-genotype relationships in monogenic disease: lessons from the thalassaemias.* Nat Rev Genet, 2001. **2**(4): p. 245-55.
176. Weatherall, D.J., et al., *A form of hereditary persistence of fetal haemoglobin characterized by uneven cellular distribution of haemoglobin F and the production of haemoglobins A and A2 in homozygotes.* Br J Haematol, 1975. **29**(2): p. 205-20.
177. Weatherall, D.J. and J.B. Clegg, *Inherited haemoglobin disorders: an increasing global health problem.* Bull World Health Organ, 2001. **79**(8): p. 704-12.
178. Weatherall, D.J., J.B. Clegg, and M.A. Naughton, *Globin synthesis in thalassaemia: an in vitro study.* Nature, 1965. **208**(5015): p. 1061-5.
179. Westin, E.H., et al., *Expression of cellular homologues of retroviral onc genes in human hematopoietic cells.* Proc Natl Acad Sci U S A, 1982. **79**(8): p. 2490-4.
180. Wijgerde, M., et al., *The role of EKLF in human beta-globin gene competition.* Genes Dev, 1996. **10**(22): p. 2894-902.

181. Wijgerde, M., F. Grosveld, and P. Fraser, *Transcription complex stability and chromatin dynamics in vivo.* Nature, 1995. **377**(6546): p. 209-13.
182. Wilber, A., A.W. Nienhuis, and D.A. Persons, *Transcriptional regulation of fetal to adult hemoglobin switching: new therapeutic opportunities.* Blood. **117**(15): p. 3945-53.
183. Wood, W.G., et al., *The determination of free and protein-bound haemoglobin in plasma using a combination of HPLC and absorption spectrometry.* Clin Lab, 2001. **47**(5-6): p. 279-88.
184. Xiao, C., et al., *MiR-150 controls B cell differentiation by targeting the transcription factor c-Myb.* Cell, 2007. **131**(1): p. 146-59.
185. Zhang, X., Q. Xu, and I. Saiki, *Quercetin inhibits the invasion and mobility of murine melanoma B16-BL6 cells through inducing apoptosis via decreasing Bcl-2 expression.* Clin Exp Metastasis, 2000. **18**(5): p. 415-21.
186. Zhou, D., et al., *KLF1 regulates BCL11A expression and gamma- to beta-globin gene switching.* Nat Genet, 2010. **42**(9): p. 742-4.
187. Zuber, J., et al., *An integrated approach to dissecting oncogene addiction implicates a Myb-coordinated self-renewal program as essential for leukemia maintenance.* Genes Dev, 2011. **25**(15): p. 1628-40.

10 APPENDICES

Appendix 1: Genotyping Results

Table 1.1 *HMIP* block 2, *MYB* SNP rs210796 and *HBS1L* SNP rs13064 Genotypes

No	HbF # ID	HbF	F Cells %	*HMIP* block 2	MYB rs210796 A/T	HBS1L Exon 17 rs13064(A/G)
1	6	0.3	4.2	(+/-) genotype	A/T	A/G
2	7	0.5	10	(+/-) genotype	A/A	-
3	9	0.5	9	(+/-) genotype	T/T	A/G
4	13	0.3	6	(+/-) genotype	A/A	A/G
5	14	0.2	3	(+/-) genotype	A/T	A/G
6	15	0.4	6	(+/-) genotype	A/T	A/G
7	20	0.3	4.6	(+/-) genotype	A/A	-
8	21	0.3	5	(+/-) genotype	A/T	-
9	31	1	13	(+/-) genotype	A/T	-
10	34	0.2	4	(+/-) genotype	A/T	-
11	35	0.3	7	(+/-) genotype	A/T	-
12	38	0.2	4	(+/-) genotype	A/T	-
13	41	0.4	7	(+/-) genotype	A/A	A/G
14	45	0.4	7	(+/-) genotype	A/A	-
15	48	0.4	6	(+/-) genotype	A/A	-
16	51	0.2	4	(+/-) genotype	A/A	-
17	19	0.2	4	(+/-) genotype	A/A	-
18	8	0.1	2	(-/-) genotype	T/T	A/G
19	11	0.2	4.5	(-/-) genotype	T/T	A/G
20	18	0.2	3.5	(-/-) genotype	A/T	-
21	22	0.2	2.5	(-/-) genotype	A/A	-
22	23	0.3	3.5	(-/-) genotype	A/T	-
23	24	0.2	3	(-/-) genotype	A/T	-
24	25	0.2	3.5	(-/-) genotype	A/A	-
25	26	0.6	9	(-/-) genotype	A/T	-
26	27	0.4	7	(-/-) genotype	A/A	-
27	28	0.3	6	(-/-) genotype	A/T	-
28	29	0.2	5	(-/-) genotype	A/T	A/G
29	30	0.4	7	(-/-) genotype	A/T	-
30	33	0.2	7	(-/-) genotype	T/T	-
31	36	0.2	3	(-/-) genotype	A/A	-
32	39	0.3	5	(-/-) genotype	A/A	-
33	40	0.3	4	(-/-) genotype	A/T	A/A
34	42	0.2	4.5	(-/-) genotype	A/A	A/G
35	43	0.3	6	(-/-) genotype	A/A	-
36	44	0.2	5	(-/-) genotype	A/A	-
37	47	0.4	5	(-/-) genotype	A/T	A/G
38	49	0.3	4	(-/-) genotype	A/T	-
39	16	0.2	3	(-/-) genotype	A/T	-
42	12	1	15	(+/+) genotype	A/T	A/A
41	46	1	15	(+/+) genotype	A/A	-
42	1	1	15	(+/+) genotype	T/T	A/A

Table 2.1 EMSA Oligonucleotides

Oligonucleotide	**Sequence**	**Encompassed SNPs**
Wt oligo	5^{I-}GATGTTACTATATCAAAACCACAA3^{I} 3^{I} CTACAATGATATAGTTTTGGTGTT5^{I-}	rs66650371(TAC/---)
Del oligo	5^{I-}AGATGTTATATCAAAACCACAAAA3^{I} 3^{I} TCTACAATATAGTTTTGGTGTTTT5^{I-}	rs66650371(TAC/---)
Old-A oligo	5^{I-}CAGGCGCATGCAACCACACCCGACT-3^{I} 3^{I}-GTCCGCGTACGTTGGTGTGGGCTGA-5^{I}	Rs9389268 (G/A) Rs9376091 (T/C)
Old-G oligo	5^{I}-CAGGCGCATGCAACCGCACCTGACT-3^{I} 3^{I}-GTCCGCGTACGTTGGCGTGGACTGA-5^{I}	Rs9389268 (G/A) Rs9376091 (T/C)
Rs9494142 T oligo	5^{I} ACTACTGACATTTATCAACATGGTG3^{I} 3^{I}TGATGACTGTAAATAGTTGTACCAC5^{I}	Rs9494142(T/C)
Rs9494142 C oligo	5^{I} ACTACTGACACTTATCAACATGGTG3^{I} 3^{I}TGATGACTGTGAATAGTTGTACCAC5^{I}	Rs9494142(T/C)

Table 2.2: Snapshot Extension Primers

Extension Primer	**Primer Sequence**	**Encompassed SNP**
MYB intron4b ext F	5[I] ATTTGAGGAAGCAGGTAAA3[I]	Rs210796(A/T)
HBS1Lexon17ext F16	5[I] ACCTCCCCAGCTCTTT3[I]	rs13064(A/G)
Rs9389268 ext F	5[I] ACAGGCGCATGCAACC3[I]	Rs9389268(G/A)
Rs9494142 ext F	5[I] ATTCTACTACTGACA3[I]	Rs9494142(T/C)

Table 2.3 ChIP-qPCR primers for GATA-1 ChIP

Primer target	Forward primer sequence	Reverse primer sequence	Amplicon Size
α-HS40 Positive control for GATA1 binding	5' CGACCCTCTGGAACCTATCAG3'	5' TTGGCCTCCAGAAGCACTG3'	151bp
β-HS 2 Positive control for GATA1 binding	5GGGAGCAGTCCCATGTAGTAGT3'	5TAGGTGGTTAGGTCAGGTTGGT3'	153bp
NEFM Housekeeping gene, negative control and normaliser	5' AGCTGGACAAGAAGGTGCAG3'	5' CCGTCGAGATGTCTGTCTTC3'	150bp
Gata-1 Peak 1	5' TGGCATCACCTGGCCTCATAA3'	5' TGGCTATCAGATCAAGTCACCAAA3'	105bp
Gata-1 Peak 2	5' CTCCATGGACAGCTGCCACTTTAA3'	5' TCCAGTGCCTCTGCTCCCATCC3'	118bp
Gata-1 Peak 3	5' AAAGCACACCTCTAACTGATGGAA3'	5' TCCGGTTACCATAGAGGAGAAGAA3'	111bp
Gata-1 Peak 3-2	5' TGGTTTTAGGCTCGGTTGTGAAAA3'	5' GTAAGTGTCTTCTGAGGGAACCG3'	181bp
Gata-1 Peak 4	5' AGGCTGCTGGCTTCTTTGCT3'	5' TCAACATGGTGGGTGTGATATCTT3'	158bp
Gata-1 Peak 5	5' TCACCTCACCACACCACTGCATA3'	5' TGACAGGTTGTTCCAACCCATTGT3'	161bp

CTSG *Cathepsin-G* promoter Positive control for MYB binding	5TTGATGGGTTGAGGGTTCTGG3[I]	5CAGTCAGTTGCTGCTGTGCTTC3[I]	304bp
HBB promoter (*β-globin*) Positive control for KLF1 binding	5[I] CAATAGATGGCTCTGCCCTGAC3[I]	5[I] CAGGTACGGCTGTCATCACTTA3[I]	129bp
NEFM Housekeeping gene, negative control and normaliser	5[I] AGCTGGACAAGAAGGTGCAG3[I]	5[I] CCGTCGAGATGTCTGTCTTC3[I]	150bp
MYB / KLF ChIP primer for rs9399137	5[I] AACATCACCTTAAAAGGCGGT	5[I] TCCACTTTCAGAACTTATCCCAAG3[I]	147bp
MYB / KLF ChIP primer for rs9389268	5[I] TGGGTAGCTGAGATTACAGGCGCAT	5[I] ACTGAGGCAGGTGGATTGCCTGAGC3[I]	131bp
MYB / KLF ChIP primer for rs55936352	5[I] GGAGGGGGTCAGTTGAAATTTTAGG	5[I] GTCAGCGTGATCAGGATATACAGCA3[I]	147bp

Table 2.4 ChIP-qPCR primer sequences for MYB and KLF1 ChIP

Table 2.5: Allele Specific Expression Primers

Primer Targets	**Forward primer sequence**	**Reverse primer sequence**	**Amplicon Size**
HBS1L exon17	5[I] CGTAGCATGAACCTCCCCAG3[I]	5[I] GCCAGAATGCATTGGTAGAGC3[I]	91bp
MYB	5[I] ACCGTATTTCTGTACAAGCTCTA3[I]	5[I] ATTATGAGTCACACACCAATGAAGA3[I]	260bp

Intron 4b			

Table 2.6 Primers and Probes used for Gene Expression Profiling

Gene	Forward primer sequence	Reverse primer sequence	Probe sequence
MYB	5' ATGATGAAGACCATGAGAAGGAAA3'	5' AACAGGTGCACTGTCTCCATGA3'	5' TGCTACCAACACAGAACCACACATGCA3'
HBS1L	5' ACAAGAATGAGGCAACAGTATCTACAG3'	5' TCAGATTCACTTCGCGATGTCT3'	5' CAAAAGGAAAACCAGTAGATTCC3'
GATA-1	5' CAATGCCTGCGGCCTCTA3'	5' CATCCTTCCGCATGGTCAGT3'	5' CTACAAGCTACACCAGGTGA3'
KLF1	5' CCCAGTCACTAGGAGAGTCCAA3'	5' GCTACACCAAGAGCTCCCACCT3'	No probe (SYBR Green detection)

Table 2.7 List of Antibodies

1º antibody to	Source	Molecular

target protein		weight (kDa) of target protein
Actin	Abcam	40
MYB	Millipore 1:1	75
KLF1	Provided by Prof. Sjaak Philipsen	38
GATA-1	Santa Cruz M-20 sc 1234x	45

Appendix 3: ChIP results

Figure 3.1: GATA-1 ChIP (not normalized)

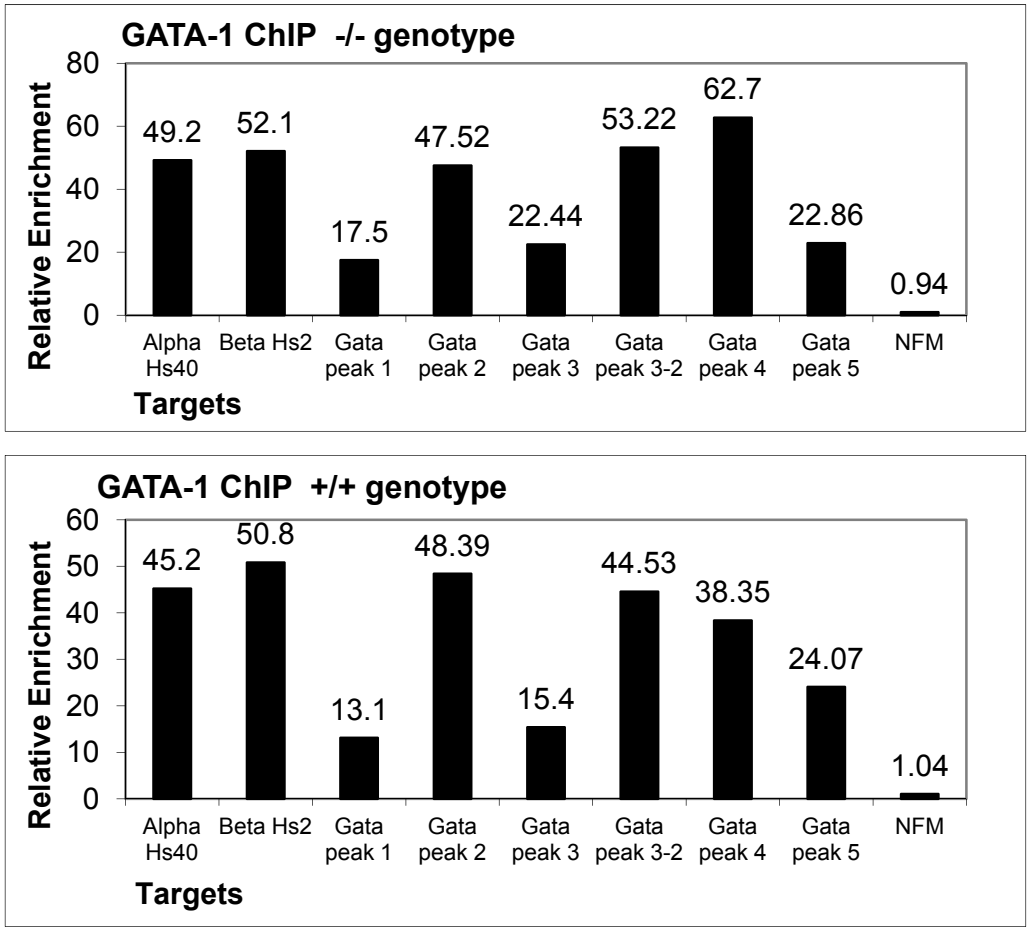
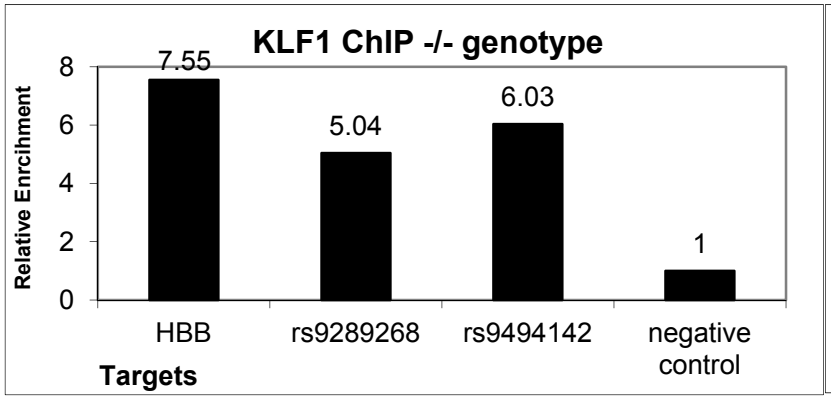


Figure 3.2: KLF1 ChIP (not normalized)



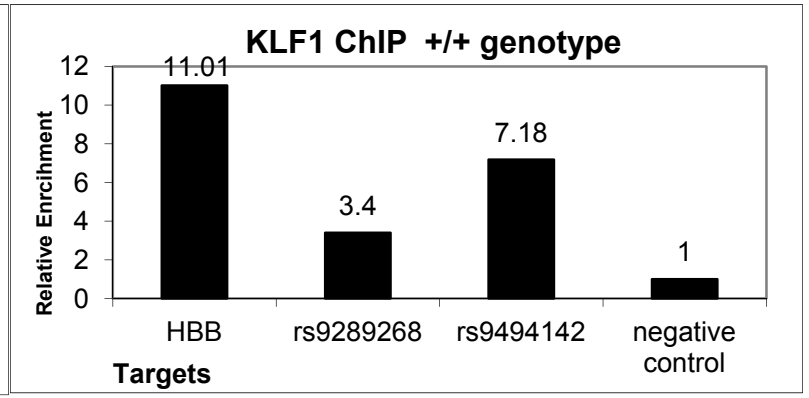
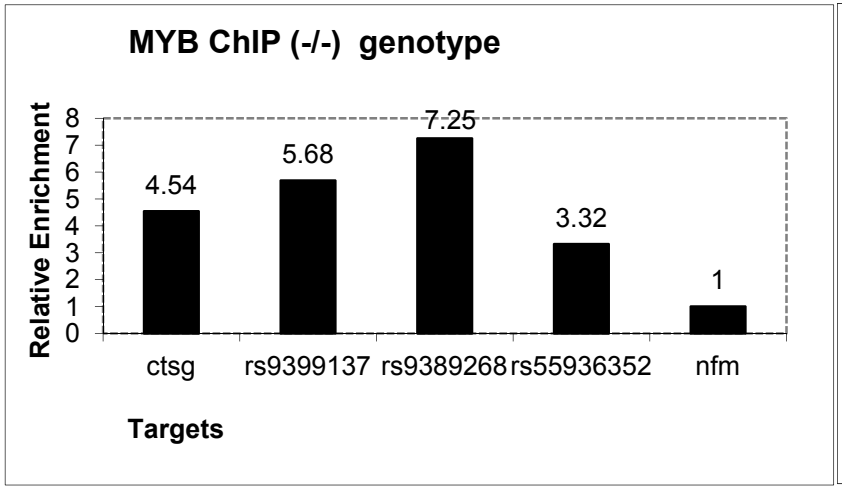
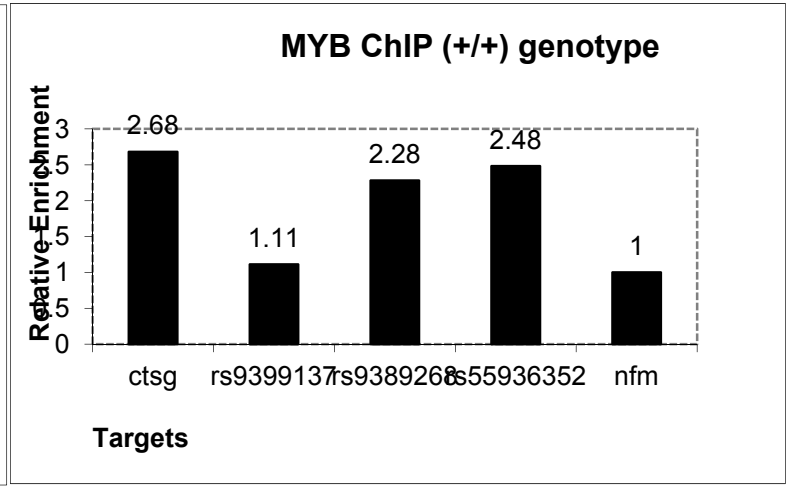


Figure 3.3: MYB ChIP (not normalized)





Appendix 4: Snapshot Results (Raw Data)

Figure 4.1: Snapshot for *MYB* SNP rs210796 (A/T)

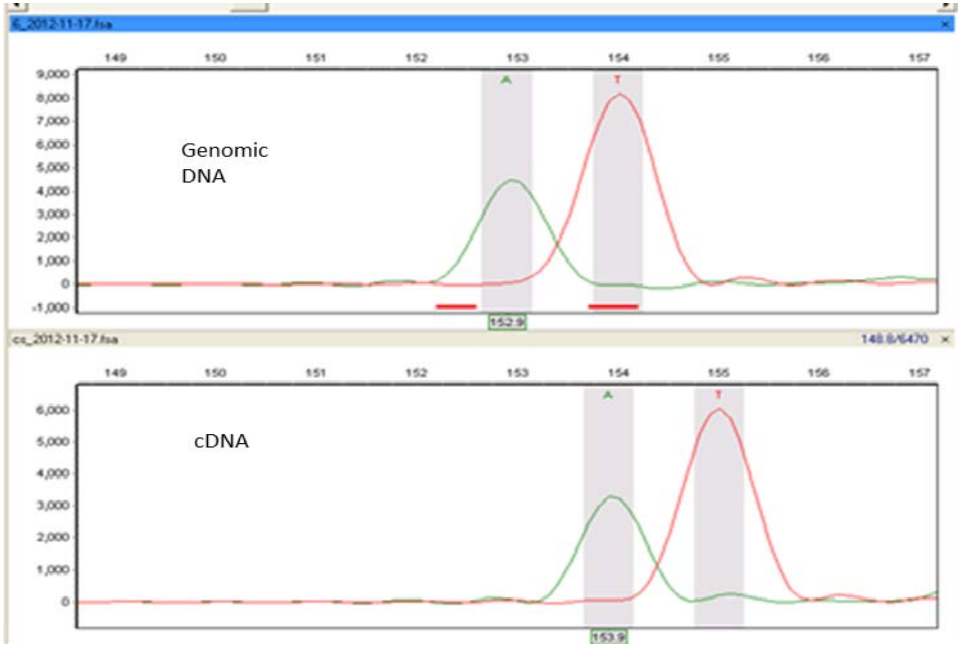
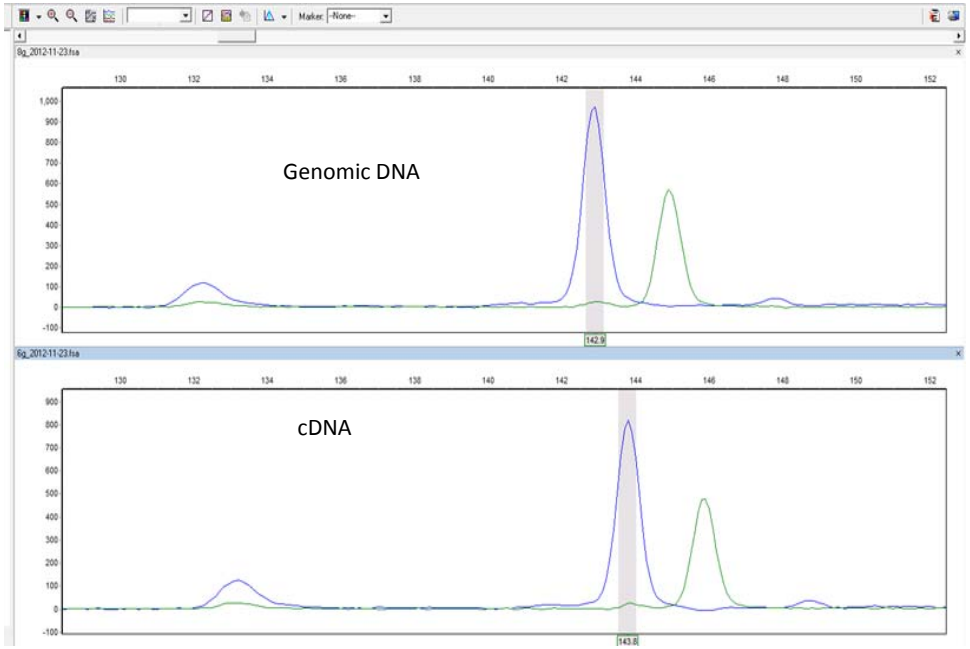


Figure 4.2: Snapshot for *HBS1L* SNP rs13064 (G/A)



Appendix 5

List of abbreviations

ACH	Active chromatin hub
AcH3	Acetylated histone H3
AcH4	Acetylated histone H4
αHS40	α-globin locus DNase I hypersensitive site 40
AML	Acute myeloid leukaemia
βHS3	β-globin LCR DNase I hypersensitive site 3
BFU-e	Burst forming unit erythroid
Bp	Base pairs
CFU-e	Colony forming unit erythroid
ChIP-chip	Chromatin immunoprecipitation on microarray
EPO	Erythropoietin
F	Fetal haemoglobin
FACS	Fluorescence activated cell sorting
FC	F-cells
FCS	Foetal calf serum
FOG-1	Friend of GATA-1
GM-CSF	Granulocyte-macrophage colony stimulating factor
GPA	Glycophorin A
HbA	Adult haemoglobin ($\alpha_2\beta_2$)
HBA	α-globin
HbA2	Haemoglobin minor ($\alpha_2\delta_2$)
HbAS	Sickle trait, heterozygous for HbS
HBB	β-globin
HBBP1	βψ-pseudogene
HBD	δ-globin
HBE	ε-globin
HbF	Foetal haemoglobin ($\alpha_2\gamma_2$)
HBG	γ-globin

HBM	ψα2-pseudogene
HBQ1	θ-globin
HbS	Sickle haemoglobin, $\alpha_2\beta^s_2$
HBZ	ζ-globin
HMIP	*HBS1L-MYB* intergenic polymorphism
HPFH	Hereditary persistence of foetal haemoglobin
HS	Hypersensitive site (to Dnase I)
HSC	Haematopoietic stem cell
IL-3	Interleukin 3
Kb	Kilo base pairs
KLF	Krüpple like factor-1
LCR	Locus control region
LD	Likage disequilibrium
Mb	Mega base pair
MYB	Myeloblastosis oncogene
NEFM	Neurofilament medium gene
PB	Peripheral blood
PBS	Phosphate buffered saline
PCR	Polymerasechainreaction
Q-PCR	Quantitative PCR
QTL	Quantitative trait locus
RNAP II	RNA polymerase II
RT-PCR	Reverse transcription PCR
SCD	Sickle cell disease
SCF	Stem cell factor
SNP	Single nucleotide polymorphism
TFBS	Transcription factor binding site
TF	Transcription factor
TSS	Transcriptional start site
UTR	Untranslated region